Graduate Texts in Mathematics

GTM

Steven P. Lalley

Random Walks on Infinite Groups

Springer

Graduate Texts in Mathematics 297

Graduate Texts in Mathematics

Graduate Texts in Mathematics bridge the gap between passive study and creative understanding, offering graduate-level introductions to advanced topics in mathematics. The volumes are carefully written as teaching aids and highlight characteristic features of the theory. Although these books are frequently used as textbooks in graduate courses, they are also suitable for individual study.

Steven P. Lalley

Random Walks on Infinite Groups

Springer

Steven P. Lalley
Department of Statistics
University of Chicago
Chicago, IL, USA

ISSN 0072-5285 ISSN 2197-5612 (electronic)
Graduate Texts in Mathematics
ISBN 978-3-031-25631-8 ISBN 978-3-031-25632-5 (eBook)
https://doi.org/10.1007/978-3-031-25632-5

Mathematics Subject Classification: 60B15, 60J50, 31CXX, 20-XX

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

In 1921, George Pólya published a short article [110] posing the following problem. Imagine a traveller on an infinite regular grid of roads—an infinite Manhattan, without Broadway—which can be viewed mathematically as the 2-dimensional integer lattice $\mathbb{Z}^2$. More generally, imagine the traveller in d-dimensional space, on the integer lattice $\mathbb{Z}^d$. The traveller starts at the origin of $\mathbb{Z}^d$; at each time $n = 1, 2, \cdots$, the traveller chooses, at random, one of the integer points at distance 1 from his/her current location, and then moves to that point, where he/she stays for one unit of time. What is the probability that the traveller will ever return to his/her original location?

Pólya proved that in dimensions $d = 1$ and $d = 2$, the probability of eventual return is 1, but that in dimensions $d \geq 3$ the probability of return is less than 1. It is not difficult to see that this implies that in dimensions 1 and 2 the traveller's trajectory is *recurrent*, that is, it will revisit the initial location infinitely often, with probability one, but in dimensions 3 and higher it is *transient*, that is, it will eventually wander away from the initial point, never to return. The notion of recurrence in a dynamical system had been introduced some years earlier by H. Poincaré, but it was Pólya's article that made it a central theme in the study of random processes.

Pólya's random walks have analogues in other finitely generated groups. For a group Γ with a finite, symmetric generating set $\mathbb{A}$, form the *Cayley graph* G by putting edges between those pairs g, g' of group elements such that $g' = ga$ for some element $a \in \mathbb{A}$. Given a probability distribution μ on the generating set $\mathbb{A}$ and an infinite sequence of $\mathbb{A}$-valued random variables all with distribution μ, one may execute a random walk on Γ by using the sequence to determine an infinite sequence of steps across edges of G.

When are such random walks recurrent? transient? If a random walk is transient, where does it go? and how fast does it get there? Are properties like transience and recurrence completely determined by the group? Answers have emerged to many—but not all—of these questions over the past 70 years, and as they have done so the study of random walks has become an important tool for understanding certain structural properties of the groups in which they take place.

In the summer of 2018, I gave a 1-week series of lectures on random walks to an audience of advanced undergraduate and beginning graduate students in mathematics at the Northwestern University Probability Summer School, organized by Antonio Auffinger and Elton Hsu. This book is an outgrowth of the lecture notes that I prepared for the course. Since the course consisted of only 6 one-hour lectures, its objectives were necessarily limited: the high points were (i) a discussion of speed, entropy, and spectral radius, with a sketch of the subadditive ergodic theorem; (ii) the Carne-Varopoulos inequality and its implications; (iii) Kesten's theorem on nonamenability and the exponential decay of return probabilities; and (iv) the equivalence of positive speed, positive entropy, and the existence of nonconstant, bounded harmonic functions. This book covers quite a bit more, but like the course its scope is limited. It is not meant to be a comprehensive treatise, but rather an accessible introduction to the subject that can be profitably read by graduate students in mathematics who have a working knowledge of Lebesgue measure and integration theory at the level of (say) Royden [111], and elementary group theory and linear algebra at the level of Herstein [65]. In particular, the book has been organized and written so as to be accessible not only to students in probability theory, but also to students whose primary interests are in geometry, ergodic theory, or geometric group theory. For this reason, some important subjects in elementary probability theory—the Ergodic Theorem, the Martingale Convergence Theorems, and several others— have been integrated into the exposition. I have also included an Appendix with an introduction to the terminology, notation, and basic notions of measure-theoretic probability. I have not attempted to compile a comprehensive bibliography; instead, I have limited the bibliography to articles either directly related to topics discussed in the text, or on related subjects that I judged might be of interest to beginning researchers.

Reader's Guide. The first seven chapters form the core of the book; they could be used as the basis of a short course in the subject. These chapters deal with the most basic properties of random walks: transience/recurrence, speed, entropy, decay of return probabilities, and the relation between harmonic functions and hitting probabilities. They also introduce some of the most important tools of the trade: the ergodic theorems of Birkhoff and Kingman, Markov operators, Dirichlet forms, and isoperimetric inequalities. Finally, they culminate, in Chapter 7, in one of the subject's landmark achievements, Varopoulos' classification of recurrent groups.

Chapters 8–10 are devoted to the study of bounded harmonic functions. Among the highlights of the theory are Blackwell's characterization of such functions (Chapter 9) and the theorem of Avez, Derriennic, and Kaimanovich & Vershik showing that a random walk has the Liouville property if and only if it has Avez entropy 0 (Chapter 10). In Chapter 11, I discuss group actions and their use in the study of random walks. Here I introduce the important notion of a *boundary*, and show how one important special case, the *Busemann boundary*, figures in Karlsson and Ledrappier's characterization of speed. Chapter 12 lays out the basic facts about *Poisson boundaries*, including Kaimanovich and Vershik's entropic characterization of Poisson boundaries.

The last three chapters deal with several more specialized topics: hyperbolic groups (Chapter 13); unbounded harmonic functions (Chapter 14), and (iii) Gromov's classification of groups of polynomial growth (Chapter 15), a special case of which is used in the proof of Varopoulos' growth criterion for recurrent groups in Chapter 7. The latter two chapters are independent of Chapters 8–13; Chapter 15 depends logically on the main result of Chapter 14, but could be read separately.

I have included a large number of exercises, especially in the early chapters. These are an integral part of the exposition. Some of the exercises develop interesting examples, and some fill crevices in the theory; many of the results developed in the exercises are used later in the book. Those exercises that can be skipped at first reading are marked with a dagger †, as are several sections of the text not used elsewhere.

Acknowledgment Professor Si Tang of Lehigh University read the entire first draft of the manuscript. This final version has benefitted greatly from her comments and suggestions. Professor Omer Tamuz of Caltech also read parts of the manuscript and made a number of useful suggestions (one of which was to get rid of an egregious error in Chapter 4—thanks for that, Omer!).

Chicago, IL, USA Steven P. Lalley

Contents

Chapter 1
First Steps

1.1 Prologue: Gambler's Ruin

I have A dollars; my friend Si has B dollars. A cup of coffee at the Sacred Grounds Coffee Shop in Swift Hall costs $A + B$ dollars, which neither Si nor I can afford, so we decide to do the only sensible thing: we gamble for it. To make the game fair, we agree to each bet 1 dollar at a time, with the winner decided by a fair coin toss; if the coin comes up Heads, Si wins my dollar, but if it comes up Tails then I win her dollar. We play this game repeatedly, until one or the other of us has won all $A + B$ dollars.

Problem: What is the probability that I win?

This problem can be reformulated as a first-exit problem for the *simple random walk* on the integers, as follows. Let $X_1, X_2, X_3, \dots$ be an infinite sequence of independent random variables, each with distribution

$$P\{X_i = +1\} = P\{X_i = -1\} = \frac{1}{2}, \tag{1.1.1}$$

and for each $n = 1, 2, 3, \dots$ set

$$S_n = \sum_{i=1}^{n} X_i. \tag{1.1.2}$$

The sequence $(S_n)_{n \geq 0}$ is the *simple random walk* on the group $\mathbb{Z}$; at each step, the random walker moves either one space to the right or one to the left, with equal probabilities for the two possibilities. The problem of immediate interest is to determine the probability that the random walk will visit the site $+B$ before it visits $-A$.

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_1

Exercise 1.1.1 Show that the event that the random walk never exits the interval $(-A, +B)$ has probability 0.

HINT: If there is ever a run of $A + B$ consecutive $+1$s in the sequence $X_1, X_2, X_3, \ldots$ then the random walk will exit $(-A, +B)$, if it hasn't already done so.

To solve our problem, let's consider the superficially more general scenario in which the random walker starts at an arbitrary integer $x \in [-A, B]$. This can be arranged by setting $S_n^x = x + S_n$; the event G that the random walk $(S_n^x)_{n\geq 0}$ visits $+B$ before $-A$ is identical to the event that the random walk $(S_n)_{n\geq 0}$ visits $B - x$ before $-A - x$. Define $u(x)$ to be the probability of this event. Clearly, the function $u(x)$ satisfies the "boundary conditions"

$$u(B) = 1 \quad \text{and} \quad u(-A) = 0, \tag{1.1.3}$$

because if the random walk starts at one of the two endpoints then there is no chance that it will visit the other endpoint first. For any other starting position $x \in (-A, B)$, the random walker must make at least one step before reaching one of the endpoints, and the first step must be to either $x - 1$ or $x + 1$. Because all subsequent steps are independent of the first step, the future of the random walk is that of a new random walk, started either at $x + 1$ or $x - 1$, with increments $X_2, X_3, \cdots$. Conditional on the event that this new random walk is started at $x + 1$, the probability that it will reach B before $-A$ is $u(x + 1)$; if it is started at $x - 1$, then the probability is $u(x - 1)$. Thus, the function u satisfies the *mean value property*: for every integer $x \in [-A + 1, +B - 1]$, the value $u(x)$ is the mean of the values $u(x \pm 1)$ at the two neighboring sites:

$$\begin{aligned} u(x) &= P(\{S_1^x = x + 1\} \cap G) + P(\{S_1^x = x + 1\} \cap G) \\ &= P\left\{S_1^x = x + 1\right\} u(x + 1) + P\left\{S_1^x = x - 1\right\} u(x - 1) \\ &= \frac{1}{2} u(x + 1) + \frac{1}{2} u(x - 1). \end{aligned} \tag{1.1.4}$$

This equation implies that the successive differences $u(x + 1) - u(x)$ are all equal, and so it follows that the function $x \mapsto u(x)$ is linear on $[-A, B]$, with boundary values (1.1.3). There is only one such function:

$$u(x) = \frac{A + x}{A + B} \quad \text{for all integers } \; x \in [-A, B]. \tag{1.1.5}$$

The formula (1.1.5) has an interesting consequence. Fix a positive integer B, and consider the event that the simple random walk *ever* visits the site B. This event is at least as likely as the event that the random walk visits B before $-A$, where $-A$ is any fixed negative integer; consequently, by (1.1.5), the probability that the simple random walk ever visits B is at least $A/(A + B)$. But A can be taken arbitrarily

large, so it follows that

$$P\{S_n = B \ \text{ for some } n \in \mathbb{N}\} = 1.$$

The same argument, with the roles of $-A$ and $+B$ reversed, shows that for any negative integer $-A$,

$$P\{S_n = -A \ \text{ for some } n \in \mathbb{N}\} = 1.$$

Thus, simple random walk on the integers is *recurrent*: it must, with probability 1, visit every integer. This in turn implies that it must visit every integer *infinitely often*. (Exercise: Explain why.) This fact was first observed by Pólya [110], using a different argument which we will reproduce later in this chapter. We will also see (in Exercise 1.7.7) that the *expected* time of first return to the initial point is infinite.

The argument used to obtain the hitting probability formula (1.1.5) relies on the fact that the ambient group is the integers. However, it illustrates an important point that will become one of the major themes of this book: hitting probabilities for random walks are intimately connected to *harmonic functions*, that is, functions which satisfy mean value properties.

1.2 Groups and Their Cayley Graphs

The random walks we will study in these lectures all live in infinite, finitely generated groups. We assume that the reader is familiar with the basic notions of group theory — see, for instance, Herstein [65], Chapter 1. Let's begin by fixing some terminology and identifying a few interesting examples.

Definition 1.2.1 A group Γ is said to be *finitely generated* if there is a finite set $\mathbb{A}$, called a *generating set*, such that every element of Γ is a finite product of elements of $\mathbb{A}$.

Because the set of finite sequences with entries in a finite set is countable, every finitely generated group is either finite or countably infinite. Although there are many interesting and important questions about random walks on finite groups, they will not concern us in this book: we will only consider *infinite*, finitely generated groups.

Assumption 1.2.2 *Assume throughout this book — unless otherwise explicitly stated — that* Γ *is an infinite, finitely generated group. Assume also that the chosen set* $\mathbb{A}$ *of generators is finite, does not include the group identity* 1*, and is* symmetric*, that is,* $a \in \mathbb{A}$ *if and only if* $a^{-1} \in \mathbb{A}$*.*

A finitely generated group, although by definition a purely algebraic entity, has an associated geometry which, as we will see, determines many of the qualitative properties of random walks in the group. This geometry is encapsulated by the so-

called *word metric* $d = d_{\mathbb{A}}$, which is defined as follows:

$$d(x, y) = d_{\mathbb{A}}(x, y) = \min \left\{ m \geq 0 \,:\, x^{-1}y = a_1 a_2 \cdots a_m \right\} \tag{1.2.1}$$

where the min is taken over all finite words $a_1 a_2 \cdots a_m$ with entries $a_i \in \mathbb{A}$. A routine check (which you should undertake) shows that d is in fact a metric on Γ. Group elements x, y are *nearest neighbors* if $d(x, y) = 1$. Denote by

$$|x| = |x|_{\mathbb{A}} = d(1, x) \tag{1.2.2}$$

the distance to the group identity; we will refer to the functional $x \mapsto |x|$ as the *word length norm*. The (closed) ball of radius n centered at the group identity will be denoted by

$$\mathbb{B}_n = \mathbb{B}_n(1) := \{x \in \Gamma \,:\, |x| \leq n\}\,. \tag{1.2.3}$$

Since each vertex $x \in \Gamma$ has precisely $|\mathbb{A}|$ nearest neighbors (for a *set* F the notation $|F|$ indicates the *cardinality* of F), the cardinality of the ball $\mathbb{B}_n$ is bounded above by $|\mathbb{A}|^n$. For any nonempty set of vertices $U \subset V$, the *(outer) boundary* ∂U of U is defined to be the set of all vertices $v \notin U$ such that v is a nearest neighbor of a vertex $u \in U$. Thus, the outer boundary of $\mathbb{B}_n$ is the set of all group elements x of word length norm $n + 1$.

Definition 1.2.3 The *Cayley graph* $G_{\Gamma;\mathbb{A}}$ of a finitely generated group Γ with symmetric generating set $\mathbb{A}$ is the graph with vertex set Γ whose edge set consists of all unordered pairs $\{x, y\}$ of vertices such that $d(x, y) = 1$ (that is, all pairs $\{x, y\}$ such that $y = xa$ for some $a \in \mathbb{A}$).

See Bollobas [16], Chapter 1 for the basic vocabulary of graph theory. The word metric d on the vertex set Γ coincides with the usual path metric in the Cayley graph $G_{\Gamma;\mathbb{A}}$: for any two group elements x, y the word distance $d(x, y)$ is just the length (i.e., number of edges crossed) of the shortest path in $G_{\Gamma;\mathbb{A}}$ that connects x to y.

Clearly, the edge set of the Cayley graph, and hence also the word metric d and norm $|\cdot|$, depend on the choice of the generating set $\mathbb{A}$. Our notation suppresses this dependence, as the generating set will usually be fixed; if there is a danger of ambiguity, we will write $|\cdot|_{\mathbb{A}}$, $d_{\mathbb{A}}(\cdot, \cdot)$, etc. For any two finite generating sets $\mathbb{A}_1, \mathbb{A}_2$ the corresponding word length norms are comparable, in the following sense: there exist constants $0 < C_- < C_+ < \infty$ such that

$$C_-|x|_{\mathbb{A}_1} \leq |x|_{\mathbb{A}_2} \leq C_+|x|_{\mathbb{A}_1} \quad \text{for all } x \in \Gamma. \tag{1.2.4}$$

Definition 1.2.4 An *action* of a group Γ on a set Y is a group homomorphism Φ from Γ to the group of bijections of Y. The bijection $\Phi(g)$ associated with a group element g is usually denoted by $y \mapsto g \cdot y$ or $y \mapsto gy$. Thus,

$$\begin{aligned} 1 \cdot y &= y && \text{for all } y \in Y \text{ and} \\ (gh) \cdot y &= g \cdot (h \cdot y) && \text{for all } y \in Y \text{ and} \quad g, h \in \Gamma. \end{aligned} \tag{1.2.5}$$

For any point $y \in Y$, the set $\Gamma \cdot y = \{g \cdot y : g \in \Gamma\}$ is called the *orbit* of y under the group action. If the set Y is a topological space, then we say that Γ *acts by homeomorphisms* on Y if the bijections $\Phi(g)$ are homeomorphisms; similarly, if Y is a metric space then Γ *acts by isometries* on Y if the bijections $\Phi(g)$ are isometries of Y.

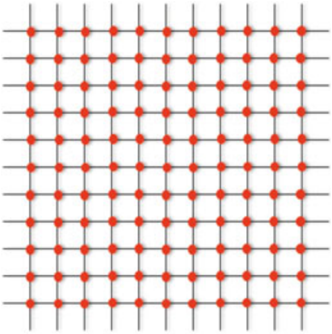
A finitely generated group Γ acts on its Cayley graph $G_{\Gamma;\mathbb{A}}$ in a natural way by left translation, sending vertices to vertices by the rule $\Phi(g)(x) = gx$, where gx is the group product. This action on vertices also maps edges to edges, because if vertices a, b are nearest neighbors then so are ga, gb. Moreover, for each $g \in \Gamma$ the mapping $\Phi(g) : G_{\Gamma;\mathbb{A}} \to G_{\Gamma;\mathbb{A}}$ is an isometry (relative to the word metric); equivalently, the word metric is *invariant* (or more precisely, *left invariant*) under the action of Γ. In general, a metric on a group Γ is said to be *invariant* if

$$d(x, xy) = d(1, y) \quad \text{for all } x, y \in \Gamma. \tag{1.2.6}$$

Exercise 1.2.5 A sequence $\gamma = (x_n)_{n \in I}$ (finite or infinite) of distinct group elements indexed by an interval I of the integers $\mathbb{Z}$ is called a *geodesic path* if for any two points x_m, x_n on the path, $d(x_m, x_n) = |n - m|$, that is, if any subpath of γ is a shortest path between its endpoints. Show that in any infinite, finitely generated group Γ there is a doubly-infinite geodesic path $\cdots, x_{-1}, x_0, x_1, \cdots$ such that $x_0 = 1$.

HINTS: (A) There are group elements g at arbitrarily large distances from the identity 1; consequently, there exist geodesic segments of arbitrarily large lengths. (B) The word metric is invariant, so by translation there are geodesic segments passing through 1 whose endpoints are both at large distances from 1. (C) Use the fact that the group is finitely generated to show that there is a sequence of such geodesic segments through 1 that converge to a doubly infinite geodesic.

Example 1.2.6 The additive group $\mathbb{Z}^d$ of d-dimensional integer points has generating set $\{\pm e_i\}_{1 \le i \le d}$, where e_i is the ith standard unit vector in $\mathbb{R}^d$. The Cayley graph with respect to this set of generators is the cubic lattice in d dimensions. The group identity is the zero vector $\mathbf{0}$. The word metric is *not* the usual Euclidean metric, but rather the ℓ^1, or *taxicab*, metric.



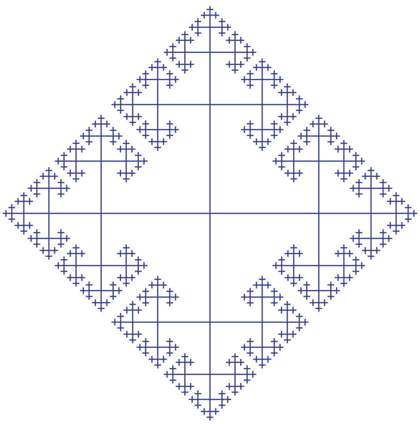
Example 1.2.7 The *free group* $\mathbb{F}_d$ on $d \geq 2$ generators $a_1, a_2, \cdots, a_d$ is the set of all finite words (including the empty word $1 = \emptyset$, which represents the group identity)

$$x = b_1 b_2 \cdots b_n$$

in which each b_i is one of the generators a_j or its inverse a_j^{-1}, and in which no entry b_i is adjacent to its inverse b_i^{-1}. Such words are called *reduced*. Two elements $x, y \in \mathbb{F}_d$ are multiplied by concatenating their representative reduced words and then doing whatever cancellations of adjacent letters are possible at the juxtaposition point. For instance, if $x = be^{-1}fb$ and $y = b^{-1}f^{-1}bbe$ then

$$xy = (be^{-1}fb)(b^{-1}f^{-1}bbe) = be^{-1}bbe.$$

The Cayley graph of the free group $\mathbb{F}_d$ relative to the standard set of generators $\mathbb{A} = \left\{a_i^{\pm 1}\right\}_{i \leq d}$ is the infinite, homogeneous tree $\mathbb{T}_{2d}$ of degree $2d$: for $d = 2$,



In this graph, for each pair of distinct vertices x, y there is a *unique* self-avoiding path $\gamma_{x,y}$ from x to y. Any path that visits first x and then y must visit every vertex

on $\gamma_{x,y}$, and must do so in the order of appearance in $\gamma_{x,y}$. Moreover, any infinite path with initial vertex x that returns to x only finitely many times must eventually reside in one of the $2d$ rooted subtrees attached to vertex x.

Example 1.2.8 The *free product* $\Gamma = \Gamma_1 * \Gamma_2 * \cdots * \Gamma_K$ of a finite collection of groups Γ_i, all with the same identity element, is the group consisting of all reduced finite words from the alphabet $\cup_{i \leq K} \Gamma_i \setminus \{1\}$, that is, sequences $a_1 a_2 \cdots a_n$ such that

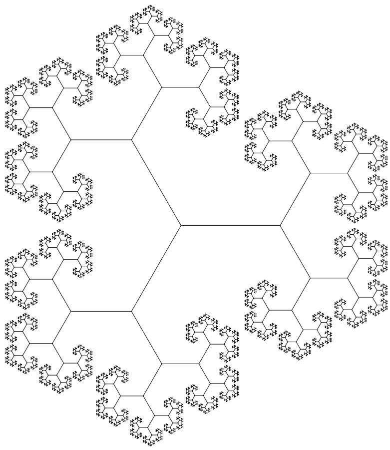
(a) each letter a_i is an element of $\Gamma_j \setminus \{1\}$ for some $j \in [K] := \{1, 2, \cdots, K\}$, and
(b) no two adjacent letters a_i, a_{i+1} are elements of the same group Γ_j.

Multiplication in the free product Γ is by concatenation followed by reduction, where reduction of a word is accomplished by successive use of the following operations:

(i) if there is an adjacent pair of letters b, b' in the same Γ_i then replace the pair bb' by their product in Γ_j; and
(ii) if this product is the group identity (i.e., if $b' = b^{-1}$) then delete the resulting 1.

The identity element is the empty word $\emptyset$. If each group Γ_i is finitely generated, then so is their free product: the union of the generating sets of the groups Γ_i is a generating set for the free product.

For instance, if $K = 3$ and each Γ_i is a copy of the two-element group $\mathbb{Z}_2$, then the free product $\mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$ can be represented by the set of all finite words (including the empty word $\emptyset$) from the three-element alphabet $\{a, b, c\}$ in which no letter a, b, or c is adjacent to itself. Group multiplication is concatenation followed by successive elimination of as many "spurs" aa, bb, or cc as possible at the juxtaposition point. (Thus, each one-letter word a, b, c is its own inverse.) The Cayley graph is once again an infinite tree, this one of degree 3:



Exercise 1.2.9 Describe the Cayley graph of the free product $\mathbb{Z}_2 * \mathbb{Z}_3$ with the natural generating set.

Subgroups of a given group Γ can have very different geometries than that of the ambient group Γ. The free group $\mathbb{F}_2$ with generators $a^{\pm 1}, b^{\pm 1}$, for instance, contains the subgroups $\langle a \rangle := \{a^n\}_{n \in \mathbb{Z}}$ and $\langle b \rangle := \{b^n\}_{n \in \mathbb{Z}}$, both of which are isomorphic copies of the integers $\mathbb{Z}$. The Cayley graphs of these subgroups look nothing like that of the free group. If a subgroup H has *finite index*, however, then the geometries of H and Γ are much more closely related: in this case, the Cayley graph of Γ consists of finitely many copies of the Cayley graph of H, with edges connecting vertices in different copies. This, as we will see, implies that random walks on a finitely generated group have many of the main behavioral features of random walks on finite-index subgroups.

Definition 1.2.10 If H is a subgroup of a group Γ then its *index* is the number of distinct right cosets $(Hx_i)_{i \leq m}$ of H in Γ. If the index of H in Γ is finite, then H is a *finite index subgroup* of Γ, and Γ is a *finite extension* of H.

If $\{Hx_i\}_{i \in S}$ are the distinct right cosets of a subgroup H in Γ, then $\left\{x_i^{-1} H\right\}_{i \in S}$ are the distinct left cosets, as you should check. Therefore, the index of a subgroup could equivalently be defined as the number of distinct left cosets.

Exercise 1.2.11 Show that if Γ is finitely generated and H is a finite index subgroup of Γ then H is finitely generated.

HINT: Let $G = G_{\Gamma;\mathbb{A}}$ be the Cayley graph of Γ relative to a finite, symmetric generating set $\mathbb{A}$.

(A) Show that there exists $D < \infty$ such that for every element $g \in \Gamma$ there is an element $h \in H$ at distance $\leq D$ from g (in the graph distance for G).
(B) Let $\mathbb{A}_H$ be the set of all elements $h \in H$ that are within distance $4D$ of the group identity (in the graph distance for G). Show that $\mathbb{A}_H$ is a generating set for H.

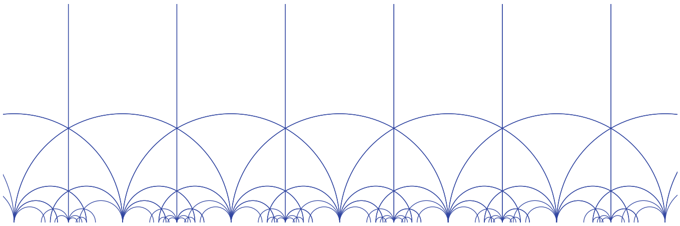
Example 1.2.12 † $SL(n, \mathbb{R})$ is the group of $n \times n$ matrices with real entries and determinant 1, with the group operation being matrix multiplication. The subgroup consisting of those elements with *integer* entries is $SL(n, \mathbb{Z})$. The group $SL(2, \mathbb{Z})$ has the two-element normal subgroup $\{\pm I\}$; the quotient group $SL(2, \mathbb{Z}) / \{\pm I\}$, known as the *modular group* $PSL(2, \mathbb{Z})$, is one of the most important groups in mathematics. The modular group acts on the upper halfplane $\mathfrak{H} = \{z = x + iy \,:\, y > 0\} \subset \mathbb{C}$ by linear fractional transformations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{az + b}{cz + d} \tag{1.2.7}$$

It can be shown (cf. [87], Chapters IV and VII) that $PSL(2, \mathbb{Z})$ has generating set

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, T^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, U = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix};$$

Its Cayley graph relative to this generating set is the *dual graph* to this:



Here each triangle represents a vertex (group element); two vertices share an edge in the Cayley graph if the corresponding triangles meet in a side.

It is by no means obvious — but nevertheless a fact — that $PSL(2, \mathbb{Z})$ is isomorphic to the free product $\mathbb{Z}_2 * \mathbb{Z}_3$. (See [2] for a short and simple proof.) Furthermore, the *commutator subgroup* of $PSL(2, \mathbb{Z})$ (the subgroup generated by all elements of the form $xyx^{-1}y^{-1}$) has finite index (=6) and is isomorphic to the free group $\mathbb{F}_2$. (See [87], Chapter XI for a full proof; altenatively, see [31], Problem II.A.16.) In Section 5.2 we will prove a weaker result, that $PSL(2, \mathbb{Z})$ has a subgroup (not necessarily of finite index) isomorphic to $\mathbb{F}_2$.

1.3 Random Walks: Definition

Definition 1.3.1 A *random walk* on a group Γ is an infinite sequence $X_0, X_1, X_2, \cdots$ of Γ-valued random variables, all defined on a common probability space $(\Omega, \mathcal{F}, P)$, such that for some nonrandom group element x (the *initial point*)

$$X_n = x\xi_1\xi_2 \cdots \xi_n \quad \text{for all } n \in \mathbb{Z}_+ := \{0, 1, 2, \cdots\} \tag{1.3.1}$$

where the random variables $\xi_1, \xi_2, \cdots$ — called the *steps* or *increments* of the random walk — are independent and identically distributed. The common distribution $\mu$of the increments ξ_i is the *step distribution* of the random walk.

If Γ is finitely generated (as we assume by default), the definition can be restated in equivalent form as follows: a sequence $(X_n)_{n\geq 0}$ of Γ-valued random variables is a random walk with initial point $X_0 = x$ if and only if there is a probability distribution μ on Γ such that for every integer $n \geq 0$ and every sequence $g_1, g_2, \cdots, g_n$ in Γ,

$$P\left(\bigcap_{m=0}^{n} \{X_m = xg_1g_2 \cdots g_m\}\right) = \prod_{i=1}^{n} \mu(g_i). \tag{1.3.2}$$

Theorem 1.3.2 *For any probability distribution μ on Γ there is a probability space $(\Omega, \mathcal{F}, P)$ that supports a random walk with step distribution μ and initial point* 1.

This is a consequence of the *Daniell-Kolmogorov Extension Theorem* of measure theory, or, more simply, the existence of Lebesgue measure on the Borel subsets of the unit interval. (See Appendix A for further explanation, along with a brief orientation to the vocabulary, notation, and conventions of probability theory.) Clearly, any probability space that supports a random walk with initial point $x = 1$ also supports random walks with arbitrary initial point $x \in \Gamma$, because if $(X_n)_{n\geq 0}$ is a random walk with initial point 1 then $(xX_n)_{n\geq 0}$ is a random walk with initial point x.

Convention: It is customary to indicate the dependence of probabilities and expectations on the initial state x of a random walk by a superscript on the probability measure; thus,

$$P^x\left(\bigcap_{m=0}^{n}\{X_m = xg_1g_2\cdots g_m\}\right) = \prod_{i=1}^{n}\mu(g_i). \tag{1.3.3}$$

When the initial state is the group identity $x = 1$, as will always be the case in these notes unless otherwise specified, the superscript will be deleted; thus, $P = P^1$. It is an easy exercise in measure theory to show that there is a measurable space $(\Omega, \mathcal{F})$ equipped with random variables $X_n : \Omega \to \Gamma$ on which probability measures P^x satisfying (1.3.3) can be simultaneously defined.

Terminology: The *support* of a step distribution μ is defined to be the set

$$\text{support}(\mu) := \{g \in \Gamma : \mu(g) > 0\}. \tag{1.3.4}$$

A random walk whose step distribution has support contained in $\mathbb{A} \cup \{1\}$ is a *nearest-neighbor random walk*; the *simple* random walk is the random walk whose step distribution is the uniform distribution on $\mathbb{A}$. A random walk on Γ (or its step distribution μ) is *symmetric* if for every $x \in \Gamma$,

$$\mu(x) = \mu(x^{-1}). \tag{1.3.5}$$

The *n-step transition probabilities* of a random walk $(X_n)_{n\geq 0}$ are defined by

$$p_n(x, y) = P^x\{X_n = y\} = P^1\left\{X_n = x^{-1}y\right\}. \tag{1.3.6}$$

For notational ease we shall usually write $p(x, y) = p_1(x, y)$. A random walk is symmetric if and only if $p(x, y) = p(y, x)$ for all $x, y \in \Gamma$. If every element y of the group is *accessible* from the initial point $X_0 = 1$, that is, if

$$\sum_{n=0}^{\infty} p_n(1, y) > 0 \quad \text{for every} \quad y \in \Gamma, \tag{1.3.7}$$

then the random walk is said to be *irreducible*.

Equation (1.3.2) and the countable additivity of probability imply that the n-step transition probability $p_n(x, y)$ can be represented as a sum of *path-probabilities* as follows:

$$p_n(x, y) = \sum_{\mathcal{P}_n(x,y)} \prod_{i=1}^{n} p_1(x_{i-1}, x_i) = \sum_{\mathcal{P}_n(x,y)} \prod_{i=1}^{n} \mu(x_{i-1}^{-1} x_i), \tag{1.3.8}$$

where $\mathcal{P}_n(x, y)$ is the set of all *paths* (i.e., sequences of group elements) $x_0, x_1, \cdots, x_n$ of length $n + 1$ from $x_0 = x$ to $x_n = y$.

Obviously, a sufficient condition for irreducibility of a nearest-neighbor random walk is that the step distribution μ attach positive probability to every element of the generating set $\mathbb{A}$. For a symmetric nearest-neighbor random walk there is no loss of generality in assuming that this latter condition holds, because if a nearest-neighbor random walk is irreducible, then the support of its step distribution must itself be a generating set.

Exercise 1.3.3 Check that the transition probabilities of a random walk satisfy the *Chapman-Kolmogorov* equations

$$p_{n+m}(x, y) = \sum_{z \in \Gamma} p_m(x, z) p_n(z, y). \tag{1.3.9}$$

Exercise 1.3.4 The *convolution* $\mu * \nu$ of two probability measures on a countable group Γ is the probability measure such that for each $z \in \Gamma$,

$$\mu * \nu(z) := \sum_{y \in \Gamma} \mu(y) \nu(y^{-1} z).$$

(A) Check that convolution is associative, that is, $(\lambda * \mu) * \nu = \lambda * (\mu * \nu)$.

(B) Show that if X and Y are independent random variables valued in Γ with distributions μ and ν, respectively, then $\mu * \nu$ is the distribution of the product XY.

Exercise 1.3.5 Show that the transition probabilities of a *symmetric* random walk satisfy the inequalities

$$p_{2n}(x, y) \leq p_{2n}(x, x) = p_{2n}(1, 1) \quad \text{and} \tag{1.3.10}$$

$$p_n(x, y)^2 \leq p_{2n}(x, x) = p_{2n}(1, 1). \tag{1.3.11}$$

HINT: Write $p_{2n}(x, y) = \sum_{z\in\Gamma} p_n(x, z)p_n(z, y) = \sum_{z\in\Gamma} p_n(x, z)p_n(y, z)$; now try every analyst's favorite trick.

Exercise 1.3.6 Show that for any transient, *symmetric* random walk on Γ and any $\varepsilon > 0$ there exists $n_\varepsilon \in \mathbb{N}$ such that for every pair of states $x, y \in \Gamma$,

$$\sum_{n=n_\varepsilon}^{\infty} p_n(x, y) < \varepsilon. \tag{1.3.12}$$

Also, show by example that this need not be true for non-symmetric random walks.

HINTS: For the first part, use the result of Exercise 1.3.5. For the second, look for a suitable random walk on $\Gamma = \mathbb{Z}$.

1.4 Recurrence and Transience

Definition 1.4.1 A random walk $(X_n)_{n\geq 0}$ on a finitely generated group Γ (or its step distribution μ) is *recurrent* if the event

$$\{X_n = 1 \text{ for some } n \geq 1\} := \bigcup_{n=1}^{\infty} \{X_n = 1\} \tag{1.4.1}$$

that the random walk eventually returns to its initial point has probability 1; otherwise, the random walk is said to be *transient* .

The distinction between recurrent and transient random walks is of natural interest. It was the problem of determining whether simple random walks on the integer lattices are recurrent or transient that initially motivated the development of the subject by Pólya in the early 1920s. Pólya [110] formulated the following criterion for determining when a random walk is recurrent.

Proposition 1.4.2 (Pólya's Criterion) *A random walk $(X_n)_{n\geq 0}$ on a finitely generated group is recurrent if and only if*

$$\sum_{n=0}^{\infty} P\{X_n = 1\} = \sum_{n=0}^{\infty} p_n(1, 1) = \infty. \tag{1.4.2}$$

Consequently, the recurrence or transience of a random walk is determined solely by the step distribution.

Proof. For each $n = 1, 2, \cdots$, let F_n be the event that the number of returns to the initial state $X_0 = 1$ is at least n. Then the total number of returns to the initial point can be written as

$$\sum_{n=1}^{\infty} \mathbf{1}\{X_n = 1\} = \sum_{n=1}^{\infty} \mathbf{1}_{F_n}.$$

Consequently, since the expectation of an infinite sum of nonnegative random variables is the sum of the expectations (by the *Monotone Convergence Theorem*),

$$\sum_{n=0}^{\infty} P\{X_n = 1\} = 1 + \sum_{n=1}^{\infty} P(F_n) = \sum_{n=0}^{\infty} P(F_1)^n = 1/(1 - P(F_1)), \qquad (1.4.3)$$

and so $P(F_1) = 1$ if and only if the sum in (1.4.2) is infinite.

The missing step in this argument is the identity $P(F_n) = P(F_1)^n$. That this holds should be intuitively clear: each time the random walk returns to its starting point, it "begins afresh", so the conditional probability that it returns again is the same as the *unconditional* probability that it returns at least once. To fashion a rigorous proof, break the event F_n into elementary pieces, that is, write it as a union of disjoint cylinder events

$$C = C(x_0, x_1, x_2, \cdots, x_m) := \bigcap_{i=0}^{m} \{X_i = x_i\} \qquad (1.4.4)$$

where $x_1, x_2, \cdots, x_m$ is a finite sequence in Γ with exactly n entries $x_i = 1$, the last at time $n = m$. For each such cylinder C, the event $F_{n+1} \cap C$ occurs if and only if C occurs and the sequence of partial products

$$\xi_{m+1}, \ \xi_{m+1}\xi_{m+2}, \cdots$$

returns to 1. Since the random variables ξ_i are independent and identically distributed, it follows that

$$P(C \cap F_{n+1}) = P(C)P(F_1).$$

Summing over the cylinder events that constitute F_n gives $P(F_{n+1}) = P(F_{n+1} \cap F_n) = P(F_1)P(F_n)$. □

Remark 1.4.3 The argument proving that $P(F_n) = P(F_1)^n$ relies on an obvious but crucial property of a random walk, known as the *Markov property*. This can be stated as follows. If $X_n = \xi_1\xi_2 \cdots \xi_n$ is a random walk with i.i.d. isteps ξ_i, then for any integer $m \geq 0$ the sequence $(X_n^*)_{n\geq 0}$ defined by

$$X_n^* = X_m^{-1} X_{n+m} = \xi_{m+1}\xi_{m+2} \cdots \xi_{m+n} \qquad (1.4.5)$$

is again a random walk with the same step distribution, and is independent of the σ-algebra $\sigma(\xi_1, \xi_2, \cdots, \xi_m)$ generated by the first m steps of the original random walk.

Definition 1.4.4 A random walk whose step distribution assigns positive probability to the group identity 1 is said to be *lazy*; the probability $\mu(1) := P\{\xi_1 = 1\}$ is called the *holding probability*.

Exercise 1.4.5 Let $(X_n)_{n\geq 0}$ be a *lazy* random walk on a finitely generated group. Show that, for any $m \in \mathbb{N} := \{1, 2, 3, ...\}$, the random walk $(X_{mn})_{n\geq 0}$ is recurrent if and only if $(X_n)_{n\geq 0}$ is recurrent.

Recurrence requires that a random walk revisits its initial point at least once almost surely. The proof of Pólya's theorem shows that if a random walk is recurrent then it must in fact revisit its initial point *infinitely often* with probability one.

Corollary 1.4.6 *If a random walk $(X_n)_{n\geq 0}$ on a finitely generated group is recurrent, then with probability* 1 *it returns to its initial point infinitely often, that is,*

$$P\left\{\sum_{n=0}^{\infty} \mathbf{1}\{X_n = 1\} = \infty\right\} = 1. \tag{1.4.6}$$

Proof. The proof of Proposition 1.4.2 shows that for each n the event F_n that the random walk makes at least n returns to its initial state has probability one. Consequently, $P(\cap_{n\geq 1} F_n) = 1$, and so the event $\cap_{n\geq 1} F_n$ that there are *infinitely* many returns has probability 1. □

Even more is true of a recurrent random walk: if a group element can be visited with positive probability, then with probability 1 it will be visited infinitely often.

Proposition 1.4.7 *Let $(X_n)_{n\geq 0}$ be a recurrent random walk on a finitely generated group Γ. For any group element $x \in \Gamma$ such that $\sum_{n=0}^{\infty} P^1\{X_n = x\} > 0$, the random walk visits the state x infinitely often with probability* 1. *More precisely,*

$$P\{N_x = \infty\} = 1, \quad \textit{where} \;\; N_x := \sum_{n=0}^{\infty} \mathbf{1}\{X_n = x\}. \tag{1.4.7}$$

We defer the proof to Section 1.8 below.

Pólya's criterion (1.4.2) shows that the transience or recurrence of a random walk $(X_n)_{n\geq 0}$ is completely determined by its *n-step transition probabilities*

$$p_n(x, y) := P^x\{X_n = y\} = P^1\left\{X_n = x^{-1}y\right\}. \tag{1.4.8}$$

This leaves open the possibility that some random walks on a group Γ might be transient, while others might be recurrent. One of our main objectives in these lectures will be to prove the following theorem of N. Varopoulos, which states

that recurrence or transience of an *irreducible, symmetric* random walk with *finitely supported* step distribution is completely determined by the geometry of the ambient group Γ.

Theorem 1.4.8 (Recurrence Type Theorem) *Every finitely generated group Γ is either* recurrent *or* transient, *in the following sense. If there is an irreducible, symmetric, recurrent random walk on Γ whose step distribution has finite support, then* every *irreducible, symmetric random walk on Γ with finitely supported step distribution is recurrent.*

In fact, we will prove somewhat more in Chapter 7: the restriction to step distributions with finite support can be relaxed to *finite second moment*.

The Recurrence Type Theorem is one of a matched pair of results relating recurrence properties of random walks to geometry. The second member of this pair is the following *Rigidity Theorem*, which asserts that the recurrence type of a finitely generated group is determined by the recurrence type of any finite index subgroup (which, by Exercise 1.2.11 must also be finite generated).

Theorem 1.4.9 (Rigidity Theorem) *A finite-index subgroup of a finitely generated group Γ has the same recurrence type as Γ.*

Like Theorem 1.4.8, Theorem 7.4.1 will be proved in Chapter 7. Here is a related but much easier fact.

Exercise 1.4.10 Let Γ and Γ' be infinite, finitely generated groups. Show that if Γ' is transient and $\varphi : \Gamma \to \Gamma'$ is a surjective homomorphism then Γ is transient.

Exercise 1.4.11 For any probability distribution μ on Γ define the *reflection* $\hat{\mu}$ of μ to be the probability distribution

$$\hat{\mu}(g) = \mu(g^{-1}). \tag{1.4.9}$$

Show that μ is transient if and only if its reflection $\hat{\mu}$ is transient.

Exercise 1.4.12 (Lazy Variants) Show that a random walk $(X_n)_{n\geq 0}$ with step distribution μ is recurrent if and only if for any $0 < r < 1$ the random walk $(X_n^*)_{n\geq 0}$ with step distribution $(1-r)\mu + r\delta_1$ is recurrent.

HINT: Without loss of generality, assume that the random walk $(X_n)_{n\geq 0}$ is defined on a probability space that also supports an independent sequence $Y_1, Y_1, \cdots$ of i.i.d. Bernoulli-r random variables. Show that the sequence $(X_n^*)_{n\geq 0}$ defined by

$$X_n^* = X_{S_n}, \quad \text{where} \quad S_n = \sum_{i=1}^{n} Y_i,$$

is a random walk with step distribution $(1-r)\mu + r\delta_1$.

Exercise 1.4.13 Prove that if $(X_n)_{n\geq 0}$ is a transient random walk on a finitely generated group Γ then with probability one,

$$\lim_{n\to\infty} |X_n| = \infty.$$

(Here $|\cdot|$ denotes the word metric norm on Γ.)

HINT: Use the fact that the group is finitely generated to show that if $|X_n| \leq R$ infinitely often for some $R < \infty$, then some group element x with $|x| \leq R$ would be visited infinitely often. Then use the identity (1.4.3) to show that for a transient random walk this is impossible.

1.5 Symmetric Random Walks on $\mathbb{Z}^d$

The *simple random walk* on the integer lattice $\mathbb{Z}^d$ is the nearest-neighbor random walk with the uniform step distribution on the natural set of generators, i.e.,

$$\mu(\pm e_i) = \frac{1}{2d} \quad \text{for } i = 1, 2, \cdots, d.$$

For random walks on $\mathbb{Z}^d$, it is customary to use additive rather than multiplicative notation for the group operation; thus, a random walk on $\mathbb{Z}^d$ is a sequence

$$S_n = \sum_{i=0}^{n} \xi_i$$

where $\xi_1, \xi_2, \cdots$ are independent, identically distributed with common distribution μ.

Pólya's Theorem *The simple random walk on $\mathbb{Z}^d$ is recurrent in dimensions $d =$ 1, 2 and transient in dimensions $d \geq 3$.*

Proof. Proof for $d = 1$. Simple random walk on the integer latice can only return to 0 at even times $2n$, and this occurs if and only if the walk makes precisely n plus steps and n minus steps. Thus, the probability of return at time $2n$ is

$$P\{S_{2n} = 1\} = \binom{2n}{n} \Big/ 2^n \sim \frac{1}{\sqrt{2\pi n}}.$$

The last approximation is by Stirling's formula $n! \sim \sqrt{2\pi n} n^n e^{-n}$; the notation $\sim$ means that the ratio of the two sides converges to 1 as $n \to \infty$. Since $\sum_{n\geq 1} n^{-1/2} = \infty$, Pólya's Recurrence Criterion implies that the random walk is recurrent. □

The higher-dimensional cases can also be handled by Stirling's formula, together with some combinatorial sleight of hand: see, for instance, Feller [39], Chapter XIV. But once we are armed with the Recurrence Type Theorem and the Rigidity Theorem of Section 1.4, we can deduce much more, with virtually no additional work.

Theorem 1.5.1 *An irreducible, symmetric random walk with finitely supported step distribution on the integer lattice $\mathbb{Z}^d$ is recurrent if $d = 1$ or $d = 2$, and transient if $d \geq 3$.*

Proof. Consider first the random walk $S_n = (S_n^{(i)})_{i \leq d}$ whose components $S_n^{(i)}$ are *independent* simple random walks on $\mathbb{Z}$. The step distribution for this random walk is

$$\mu(\pm e_1, \pm e_2, \cdots, \pm e_d) = 2^{-d}.$$

In odd dimensions d, this random walk is irreducible. In even dimensions d, the support of μ is the subgroup H of $\mathbb{Z}^d$ consisting of those vectors with even parity (i.e., whose coordinates sum to an even integer), so if the random walk is started at the origin it can only visit sites in H; however, the subgroup H has index 2, and so by the Rigidity Theorem has the same recurrence type as $\mathbb{Z}^d$. In any dimension d,

$$P\{S_{2n} = 0\} \sim \left(\frac{1}{\sqrt{2\pi n}}\right)^d,$$

since the coordinates $S_{2n}^{(i)}$ are independent, by the same calculation as in the case $d = 1$. Thus, the return probabilities are summable in dimensions $d \geq 3$, but not in dimension $d = 2$, so the random walk is recurrent if and only if $d \leq 2$. The result therefore follows from Theorems 1.4.8 and 1.4.9. □

Exercise 1.5.2 Let Γ be the group of 2×2 matrices of the form

$$M = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$$

where $a, d \in \{-1, +1\}$ and $b \in \mathbb{Z}$.

(A) Show that Γ is generated by the 6 matrices

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}, \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}.$$

(B) Check that the subset $H \subset \Gamma$ consisting of those matrices $M \in \Gamma$ with both diagonal entries equal to 1 is a finite index subgroup isomorphic to $\mathbb{Z}$.

(C) Conclude (using Theorems 1.4.8 and 1.4.9) that the random walk whose step distribution is the uniform distribution on the generating set in part (A) is recurrent.

Exercise 1.5.3 † Pólya's proof of his Recurrence Theorem for simple random walks was based on *Fourier analysis*. Following is an outline. Denote by $(S_n)_{n\geq 0}$ the simple random walk on $\mathbb{Z}^d$.

(A) Show that for any vector $\theta = (\theta_1, \theta_2, \cdots, \theta_d) \in \mathbb{R}^d$,

$$E\exp\{i\langle\theta, S_n\rangle\} = \left(\frac{1}{d}\sum_{j=1}^{d}\cos\theta_j\right)^n. \tag{1.5.1}$$

Here $\langle\cdot,\cdot\rangle$ denotes the usual inner product on $\mathbb{R}^d$.

HINT: Use the product law to show that if $S_n = \sum_{k=1}^n \xi_k$, where $\xi_1, \xi_2, \cdots$ are i.i.d., then

$$E\exp\{i\langle\theta, S_n\rangle\} = (E\exp\{i\langle\theta, \xi_1\rangle\})^n.$$

(B) Use the Fourier inversion formula to conclude that

$$P\{S_{2n} = \mathbf{0}\} = (2\pi)^{-d}\int_{\theta\in[-\pi,\pi]^d}\left(\frac{1}{d}\sum_{j=1}^{d}\cos\theta_j\right)^{2n} d\theta_1 d\theta_2\cdots d\theta_d \tag{1.5.2}$$

HINT: $E\exp\{i\langle\theta, S_n\rangle\} = \sum_{x\in\mathbb{Z}^d} P\{S_n = x\}\exp\{i\langle\theta, x\rangle\}$.

(C) Use *Laplace's method of asymptotic expansion* (see, for instance the WIKIPEDIA article on *Laplace's method*) to obtain the asymptotic formula

$$P\{S_{2n} = \mathbf{0}\} \sim 2\left(\frac{d}{4n\pi}\right)^{d/2} \quad \text{as } n \to \infty. \tag{1.5.3}$$

(D) Finally, apply Proposition 1.4.2.

1.6 Random Walks on $\mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$

Next, let's consider random walks on the free product group $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$. Recall that the Cayley graph of Γ (with respect to the natural generating set $A = \{a, b, c\}$) is the homogeneous tree $\mathbb{T}_3$ of degree 3. The vertices of this tree are identified with group elements, which in turn are finite, reduced words on the alphabet $\mathbb{A} = \{a, b, c\}$ (here the term *reduced* means that no letter appears twice in succession). The group identity 1 is the empty word; this can be viewed as the *root vertex* of the tree.

The edges of the tree $\mathbb{T}_3$ can be assigned labels from the alphabet $\mathbb{A}$ in a natural way, as follows. For any two adjacent words (=vertices) w and w', one is an

extension of the other by exactly one letter, e.g.,

$$\begin{aligned} w &= x_1 x_2 \cdots x_m \quad \text{and} \\ w' &= x_1 x_2 \cdots x_m x_{m+1}. \end{aligned}$$

For any such pairing, "paint" the edge connecting w and w' with the "color" x_{m+1}. The edge-colorings and the word representations of the vertices then complement each other, in that for any word(=vertex) $w = x_1 x_2 \cdots x_m$, the colors x_i indicate the sequence of edges crossed along the unique self-avoiding path in the tree from 1 to w.

The step distribution of a nearest-neighbor random walk on Γ is a probability distribution $\mu = \{p_a, p_b, p_c\}$ on the set $\mathbb{A}$. At each step, the random walker chooses a color at random from $\mathbb{A}$ according to the distribution μ and then moves from his/her current vertex across the edge with the chosen color. Clearly, the random walk is irreducible if and only if each $p_i > 0$.

Proposition 1.6.1 *Every irreducible, nearest neighbor random walk on the group $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$ is transient. More generally, any irreducible random walk on Γ whose step distribution has finite support is transient.*

To prove this it suffices, by the Recurrence Type Theorem, to exhibit a single transient random walk on Γ with symmetric, finitely supported step distribution. The special case of the *simple* nearest-neighbor random walk, for which the step distribution μ is the uniform distribution on $\mathbb{A}$, is the natural choice. For this random walk, we will prove a stronger result than transience, to wit, we will show that it escapes to infinity at a linear rate.

Proposition 1.6.2 *Let $(X_n)_{n \geq 0}$ be a simple random walk on Γ with initial point 1, that is, a random walk whose step distribution is the uniform distribution on the set $\mathbb{A} = \{a, b, c\}$. Then with probability 1,*

$$\lim_{n \to \infty} \frac{|X_n|}{n} = \frac{1}{3}. \tag{1.6.1}$$

Here $|X_n|$ denotes the word length norm. The limit $\lim_{n \to \infty} |X_n|/n$ is the *speed* of the random walk; we will prove later, in Chapter 3, that *every* symmetric random walk on a finitely generated group has a definite speed provided the step distribution has finite support. Positive speed implies transience, because a random walk with positive speed can make at most finitely many visits to a given state, whereas recurrence would require infinitely many visits to the initial state (cf. Corollary 1.4.6). Not every transient random walk has positive speed: for instance, Pólya's Theorem shows that simple random walk on any integer lattice $\mathbb{Z}^d$ of dimension $d \geq 3$ is transient, but these random walks have speed 0, as the following theorem implies. The problem of determining when a symmetric, nearest-neighbor random walk on a finitely generated group has positive speed will be a recurring theme of Chapters 2–5.

Proposition 1.6.2 is a consequence of Kolmogorov's *Strong Law of Large Numbers*.

Theorem 1.6.3 (Strong Law of Large Numbers) *If $Z_1, Z_2, \cdots$ are independent, identically distributed real-valued random variables such that $E|Z_1| < \infty$, then with probability* 1*,*

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Z_i = EZ_1. \tag{1.6.2}$$

The Strong Law of Large Numbers is itself a special case of Birkhoff's *Ergodic Theorem*, which we will prove in Chapter 2. (For *bounded* summands Z_i there is a much simpler proof of the SLLN, using *Hoeffding's inequality*: see Section A.6 of the Appendix. This special case is all that is needed for the examples considered in this section.) The SLLN implies that *any* symmetric random walk on $\mathbb{Z}^d$ whose step distribution μ has finite first moment $\sum_{x\in\mathbb{Z}^d} |x|\mu(\{x\})$ must have speed zero. (Exercise: Why?) Proving that the simple random walk on $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$ has speed $1/3$ is only slightly harder.

Proof of Proposition 1.6.2. We will show that on any probability space that supports a version $(X_n)_{n\geq 0}$ of the simple random walk on Γ there is a nearest-neighbor random walk $Z_n = \sum_{i=1}^{n} Y_i$ on $\mathbb{Z}$ with step distribution

$$P(Y_n = +1) = \frac{2}{3} = 1 - P(Y_n = -1) \tag{1.6.3}$$

such that

$$|X_n| - Z_n = 2N_n \geq 0 \quad \text{for every } n \in \mathbb{Z}_+, \text{ where}$$

$$N_n = \sum_{k=0}^{n-1} \mathbf{1}\{|X_k| = 0\}$$

is the number of visits to the group identity 1 before time n. The Strong Law of Large Numbers implies that $\lim_{n\to\infty} Z_n/n = EY_1 = 1/3$ almost surely, so it will follow that $\liminf_{n\to\infty} |X_n|/n \geq 1/3$, and hence that $\lim_{n\to\infty} N_n < \infty$, with probability 1; this will imply that $\lim_{n\to\infty} |X_n|/n = \lim_{n\to\infty} Z_n/n$, and hence also relation (1.6.1).

Let $\xi_n = X_{n-1}^{-1} X_n$ be the steps of the simple random walk on Γ. These are, by definition, independent, identically distributed random variables with the uniform distribution on the generating set $\mathbb{A} := \{a, b, c\}$. Define random variables Y_n by

$$\begin{aligned} Y_n &= +1 \quad \text{if } |X_n| > |X_{n-1}| \geq 1; \\ &= -1 \quad \text{if } |X_n| < |X_{n-1}| \geq 1; \end{aligned}$$

$$= +1 \quad \text{if} \quad \xi_n \neq a \quad \text{and} \quad |X_{n-1}| = 0;$$
$$= -1 \quad \text{if} \quad \xi_n = a \quad \text{and} \quad |X_{n-1}| = 0, \tag{1.6.4}$$

and let

$$Z_n = \sum_{i=1}^{n} Y_i .$$

Thus, the successive differences $|X_n| - |X_{n-1}|$ agree with the increments Y_n except when $|X_{n-1}| = 0$, where they differ by 2; hence, for any $n \geq 0$ the difference $|X_n| - Z_n$ is twice the number of visits to the group identity by the random walk $(X_j)_{j \geq 0}$ before time n.

Therefore, to finish the proof it is enough to show that the random variables Y_n are independent and identically distributed, with distribution (1.6.3). The random variable Y_n is a function of the random variables $\xi_1, \xi_2, \cdots, \xi_n$, so it will suffice to prove that for every cylinder event

$$C_n = \bigcap_{i=1}^{n} \{\xi_i = \alpha_i\}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{A}$, we have

$$P(C_n \cap \{Y_{n+1} = 1\}) = \frac{2}{3} P(C_n). \tag{1.6.5}$$

There are two cases, depending on whether or not $\alpha_1 \alpha_2 \cdots \alpha_n = 1$. If this equality holds, then on C_n the random variable X_n takes the value 1, and so $|X_n| = 0$; consequently, by (1.6.4),

$$C_n \cap \{Y_{n+1} = 1\} = C_n \cap \{\xi_{n+1} \neq a\},$$

and hence (1.6.5) holds by the hypothesis that the random variables ξ_i are i.i.d. with uniform distribution on the set $\{a, b, c\}$. If, on the other hand, $\alpha_1 \alpha_2 \cdots \alpha_n \neq 1$, then on the event C_n,

$$X_n = \beta_1 \beta_2 \cdots \beta_m$$

for some reduced word $\beta_1 \beta_2 \cdots \beta_m$ with $\beta_m \in \{a, b, c\}$. In this case,

$$C_n \cap \{Y_{n+1} = 1\} = C_n \cap \{\xi_{n+1} \neq \beta_m\},$$

and so once again (1.6.5) follows by the hypotheses on the random variables ξ_i. □

A minor variation of the proof of transience of the simple random walk on $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$ works for any finitely generated group whose Cayley graph is an infinite,

homogeneous tree $\mathbb{T}_k$ of degree $k \geq 3$. Thus, the *free group* $\mathbb{F}_d$ on $d \geq 2$ is transient; more generally, any free product

$$\Gamma = \mathbb{Z} * \mathbb{Z} * \cdots \mathbb{Z} * \mathbb{Z}_2 * \cdots \mathbb{Z}_2$$

other than $\mathbb{Z}_2 * \mathbb{Z}_2$ whose factors are copies of $\mathbb{Z}$ or $\mathbb{Z}_2$ is transient. In fact it is not difficult to show that any asymmetric, irreducible random walk on a free group has positive speed and therefore is transient. (Exercise: Check this. Hint: There is a surjective homomorphism from $\mathbb{F}_d$ to $\mathbb{Z}^d$ taking natural generators to natural generators.)

If a random walk is transient, where does it go? Can one identify a set of possible limit points toward which any random walk trajectory must, in a suitable sense, converge? In general, this question is rather slippery; formulating a suitable answer will occupy much of Chapter 12 below. But for nearest-neighbor random walks on groups whose Cayley graphs are trees, such as $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$, a complete answer can be given with no technical machinery.

For simplicity, let Γ be either a free group $\mathbb{F}_k$ or a free product $\mathbb{Z}_2 * \mathbb{Z}_2 * \cdots * \mathbb{Z}_2$ of two-element groups, and let $\mathbb{A}$ be the natural generating set (for $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$, this is $\mathbb{A} = \{a, b, c\}$). Denote by $\mathbb{T}$ the Cayley graph; thus, vertices of $\mathbb{T}$ are finite reduced words with letters in the alphabet $\mathbb{A}$. Define $\partial\mathbb{T}$,the space of *ends* of the tree, to be the set of all *infinite* reduced words

$$\omega = \alpha_1 \alpha_2 \cdots .$$

Let d be the metric on $\mathbb{T} \cup \partial\mathbb{T}$ defined by

$$d(\omega, \omega') = 2^{-n(\omega,\omega')}, \tag{1.6.6}$$

where $n(\omega, \omega') \geq 0$ is the maximal integer n such that the words ω and ω' (whether finite or infinite) agree in their first n coordinates. The topology induced by this metric is the same as the induced Euclidean topology on the tree by the embedding shown in Example 1.2.8 above.

Exercise 1.6.4

(A) Verify that d is in fact a metric on $\mathbb{T} \cup \partial\mathbb{T}$.

(B) For any finite reduced word $w = a_1 a_2 \cdots a_m$, define $\partial\mathbb{T}(w)$to be the subset of $\partial\mathbb{T}$ consisting of all infinite reduced words whose first m letters are $a_1 a_2 \cdots a_m$. Show that the sets $\partial\mathbb{T}(w)$ form a neighborhood base for the topology on $\partial\mathbb{T}$ induced by the metric d.

(C) Show that a sequence of elements $\omega^n = \alpha_1^n \alpha_2^n \cdots \in \partial\mathbb{T}$ converges to $\omega = \alpha_1 \alpha_2 \cdots \in \partial\mathbb{T}$ if and only if for every $k \in \mathbb{N}$,

$$\alpha_k^n = \alpha_k \quad \text{for all sufficiently large } n \in \mathbb{N}.$$

(D) Prove that the metric space $(\partial\mathbb{T}, d)$ is compact.

In the following exercises, let $X_n = \xi_1 \xi_2 \cdots \xi_n$ be the nearest-neighbor random walk on Γ whose step distribution μ is supported by the natural generating set $\mathbb{A}$. Assume that μ assigns positive probability to each element of $\mathbb{A}$. Define the *hitting probability function* $h : \Gamma \to [0, 1]$ by

$$\begin{aligned} h(x) &= P^1 \{X_n = x \;\text{ for some } n \geq 0\} \\ &= P^x \{X_n = 1 \;\text{ for some } n \geq 0\} \end{aligned} \tag{1.6.7}$$

Exercise 1.6.5 Show that

(A) if x has word representation $x = a_1 a_2 \cdots a_m$ then $h(x) = \prod_{i=1}^m h(a_i)$; and
(B) for each generator $i \in \mathbb{A}$,

$$h(i) = \mu(i) + \sum_{j \neq i} \mu(j) h(j^{-1}) h(i),$$

and so $0 < h(i) < 1$ for each generator i.

Thus, the hitting probabilities satisfy $0 < h(x) \leq \varrho^{|x|}$ for some constant $0 < \varrho < 1$.

Exercise 1.6.6

(A) Prove that

$$\lim_{n \to \infty} X_n = X_\infty \in \partial \mathbb{T}$$

exists with P^x-probability one for any initial point $x \in \Gamma$. (Here the convergence is with respect to the metric d.) The distribution (under P^x) of the exit point X_∞ is, sensibly enough, called the *exit distribution*, or *exit measure*, of the random walk. Denote this measure by ν_x.
(B) Let U be a nonempty open subset of $\partial \mathbb{T}$ and $(x_m)_{m \in \mathbb{N}}$ a sequence of vertices (i.e., elements of Γ) that converge to a point $\xi \in U$. Show that

$$\lim_{m \to \infty} \nu_{x_m}(U) = 1.$$

Exercise 1.6.7 For simplicity, assume in this exercise that $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$, so that each generator is its own inverse.

(A) Show that for any nonempty finite reduced word $w = a_1 a_2 \cdots a_m$,

$$P^1 \{X_\infty \in \partial \mathbb{T}(w)\} = \nu_1(\partial \mathbb{T}(w)) = \frac{h(w)}{1 + h(a_m)} = (1 + h(a_m))^{-1} \prod_{j=1}^m h(a_j).$$

(B) Conclude that

$$\sum_{i=a,b,c} \frac{h(i)}{1+h(i)} = 1.$$

(C) Using part (A) and Exercise 1.6.5, show that the exit measure ν_1 is *nonatomic*, that is, there no point $\omega \in \partial\mathbb{T}$ such that $\nu_1(\{\omega\}) > 0$.

HINT: For part (A), observe that $X_\infty \in \partial\mathbb{T}(w)$ if and only if the random walk makes an odd number of visits to the group element with word representation w.

Exercise 1.6.8 Fix a Borel subset $B \subset \partial\mathbb{T}$ and define a function $u_B : \Gamma \to [0, 1]$ by

$$u_B(x) = P^x \{X_\infty \in B\}. \tag{1.6.8}$$

(A) A function $u : \Gamma \to \mathbb{R}$ is *μ-harmonic* if for every $x \in \Gamma$,

$$u(x) = \sum_{y \in \mathbb{A}} \mu(y)u(xy). \tag{1.6.9}$$

Show that for every Borel set $B \subset \partial\mathbb{T}$ the function u_B is μ-harmonic.

(B) Show that if $\nu_1(B) > 0$ then the function u_B is strictly positive.

(C) The set of bounded μ-harmonic functions is a real vector space. Show that this vector space is infinite-dimensional.

HINT: Show that if $G_1, G_2, \cdots, G_k$ are nonempty open subsets of $\partial\mathbb{T}$ then no nontrivial linear combination $\sum_{i=1}^{k} c_i u_{G_i}$ vanishes identically on Γ. The result of Exercise 1.6.6 (B) might be useful for this.

The abundance of bounded harmonic functions for random walks on groups with d-ary trees as Cayley graphs is an important way in which these groups differ from the integer lattices $\mathbb{Z}^d$: as we will see in Chapter 9, nearest-neighbor random walks on $\mathbb{Z}^d$ have only constant harmonic functions. The relationships between random walks, bounded harmonic functions, and the geometry of the ambient group will be explored in depth in Chapters 8–12.

1.7 Lamplighter Random Walks

Imagine an infinitely long street with streetlamps regularly placed, one at each integer point. Each lamp can be either *on* or *off* (henceforth *on* will be denoted by 1 and *off* by 0); initially (at time $n = 0$) all lamps are off. A World Cup fan[1], who has perhaps been celebrating a bit too long, moves randomly from

[1] These lectures coincided with the 2018 World Cup.

streetlamp to streetlamp, randomly turning lamps on or off. His[2] behavior at each time $n = 1, 2, \cdots$ is governed by the following rules: he first tosses a fair coin twice, then

(i) if HH he moves one step to the right;
(ii) if HT he moves one step to the left;
(iii) if TH he flips the switch of the lamp at his current configuration;
(iv) if TT he does nothing.

At any time $n = 0, 1, 2, \cdots$, the state of the street is described by the pair $X_n = (S_n, L_n)$, where S_n is the position of the random walker, and $L_n \in \{0, 1\}^{\mathbb{Z}}$ describes the current configuration of the lamps. Since the lamps are initially all off, at any time n the lamp configuration L_n has all but finitely many entries equal to 0.

The lamplighter process (S_n, L_n) just described is, in fact, itself a symmetric, nearest-neighbor random walk on a finitely generated group known as the *lamplighter group* or the *(restricted) wreath product*

$$\mathbb{L}^1 = \mathbb{Z}_2 \wr \mathbb{Z}, \tag{1.7.1}$$

where $\mathbb{Z}_2 = \{0, 1\}$ is the 2-element group with addition mod 2. Elements of the wreath product $\mathbb{L}^1$ are pairs (x, ψ), where $x \in \mathbb{Z}$ and $\psi \in \oplus_{\mathbb{Z}}\mathbb{Z}_2$; here $\oplus_{\mathbb{Z}}\mathbb{Z}_2$ is the additive group of functions (or *lamp configurations*) $\psi : \mathbb{Z} \to \mathbb{Z}_2$ that take the value 0 at all but finitely many locations $y \in \mathbb{Z}$. Multiplication in $\mathbb{L}^1$ is defined by

$$(x, \psi) * (y, \varphi) = (x + y, \sigma^{-x}\varphi + \psi)$$

where the addition of configurations is entrywise modulo 2 and σ is the shift operator on configurations. A natural set of generators is the 3-element set

$$\mathbb{A} = \{(1, \mathbf{0}), (-1, \mathbf{0}), (0, \delta_0)\}, \tag{1.7.2}$$

where $\mathbf{0}$ is the configuration of all 0s and δ_y is the configuration with a single 1 at location y and 0s elsewhere. The step distribution of the random walk obeying (i)–(iv) above is the uniform distribution on the 4-element set gotten by adding the identity $(0, \mathbf{0})$ to the set of generators.

The lamplighter group is of interest in part because it (and its higher-dimensional analogues $\mathbb{L}^d := \mathbb{Z}_2 \wr \mathbb{Z}^d$) are *amenable* (see Chapter 5 below) but have *exponential growth* (see Section 3.2).

Proposition 1.7.1 *The lamplighter group* $\mathbb{L}^1$ *is* transient.

It isn't so easy to prove directly that the random walk defined by (i)–(iv) above is transient. There is, however, a simple modification of the lamplighter random walk for which transience is much easier to prove. In this random walk, the soccer

[2] Nearly all soccer hooligans are male.

hooligan behaves as follows. At each time $n = 1, 2, 3, \cdots$ he tosses a fair coin 3 times. The first toss tells him whether or not to flip the switch of the lamp at his current position x; the second toss tells him whether to then move right or left (to $x + 1$ or $x - 1$); and the third toss tells him whether or not to flip the switch of the lamp at his new location. The state of the street at time n is once again described by a pair (S_n, L_n), with S_n describing the hooligan's position and L_n the configuration of the streetlamps. This process (S_n, L_n) is no longer a nearest-neighbor random walk with respect to the natural set of generators listed above, but it is a symmetric, nearest-neighbor random walk relative to a different set of generators. (Exercise: Write out the list of generators for the modified lamplighter random walk, and convince yourself that the random walk is symmetric with respect to these generators.)

Proposition 1.7.2 *The modified lamplighter random walk (S_n, L_n) is transient.*

To prove this we will need a crude estimate of the size of the "range" of the base random walk $(S_n)_{n\geq 0}$. This sequence is a simple random walk on $\mathbb{Z}$.

Exercise 1.7.3 Define

$$\begin{aligned} S_n^+ &:= \max_{m\leq n} S_m \quad \text{and} \\ S_n^- &:= \min_{m\leq n} S_m. \end{aligned} \tag{1.7.3}$$

Prove that for each sufficiently small $\varepsilon > 0$ there exist $\delta > 0$ and $C < \infty$ such that

$$P\left\{S_n^+ - S_n^- \leq n^{\varepsilon}\right\} \leq Cn^{-1-\delta}.$$

HINT: Divide the time interval $[0, n]$ into subintervals of length $n^{2\varepsilon}$ and use the Central Limit Theorem (cf. Theorem A.7.3 of the Appendix) to estimate the probability that the change of S_j acoss any one of these three intervals is less (in absolute value) than n^{ε}.

Proof of Proposition 1.7.2. The relevance of the random variables S_n^+ and S_n^- is this: at each time the hooligan leaves a site, he randomizes the state of the lamp at that site. Thus, conditional on the trajectory $(S_m)_{m\leq n}$, the distribution of the configuration in the segment $[S_n^-, S_n^+]$ will be uniform over all possible configurations of 0s and 1s on this interval, that is, for any sample path trajectory $(s_m)_{0\leq m\leq n}$ of the lamplighter and any lamp configuration φ with support contained in $[s_n^-, s_n^+]$,

$$P\left\{(S_m)_{m\leq n} = (s_m)_{m\leq n} \text{ and } L_n = \varphi\right\} = P\left\{(S_m)_{m\leq n} = (s_m)_{m\leq n}\right\}/2^{s_n^+ - s_n^- + 1}.$$

Therefore,

$$P\{(S_n, L_n) = (0, \mathbf{0})\} \leq P\left\{S_n^+ - S_n^- \leq n^{\varepsilon}\right\} + \frac{1}{2^{n^{\varepsilon}+1}} \leq Cn^{-1-\delta} + \frac{1}{2^{n^{\varepsilon}+1}}.$$

This sequence is summable, so Pólya's criterion implies that the random walk is transient. □

Proposition 1.7.4 *Let $Z_n = (S_n, L_n)$ be the modified lamplighter random walk and $|Z_n|$ its distance from the group identity in the word length metric (relative to the standard generating set (1.7.2)). With probability 1,*

$$\lim_{n\to\infty} \frac{|Z_n|}{n} = 0. \tag{1.7.4}$$

This proposition, as we will see in Chapter 5, implies that the lamplighter group is *amenable*. Its proof will rely on the following elementary consequence of the Central Limit Theorem.

Lemma 1.7.5 *Let $S_n^{\pm}$ be as defined in equations (1.7.3), where $(S_n)_{n\geq 0}$ is the simple random walk on $\mathbb{Z}$. For any $b \geq 0$,*

$$\lim_{n\to\infty} P\left\{S_n^+ > b\sqrt{n}\right\} = \lim_{n\to\infty} P\left\{S_n^- < -b\sqrt{n}\right\} = \frac{2}{\sqrt{2\pi}} \int_b^{\infty} e^{-t^2/2}\, dt. \tag{1.7.5}$$

Before proving this, let's see how it implies Proposition 1.7.4. In Chapter 3 we will prove that for any random walk $(Z_n)_{n\geq 0}$ with finitely supported step distribution, $\lim_{n\to\infty} |Z_n|/n$ exists and is constant with probability one, so it is enough to show that the limit cannot be positive.

At any time n, only lamps at sites in the interval $[S_n^-, S_n^+]$ can be turned on, and the lamplighter must be at some site in this interval. Consequently, by making at most two complete traversals of this interval (first upward to S_n^+, then downward to S_n^-, then back up to 0), the lamplighter can visit every site with a lighted lamp, turn it off, and then return to the initial site 0 with all lights extinguished. Therefore,

$$|Z_n| \leq 4(S^{+n} - S_n^-).$$

Fix $\varepsilon > 0$ small, and let $b > 0$ be sufficiently large that

$$\frac{4}{\sqrt{2\pi}} \int_b^{\infty} e^{-t^2/2}\, dt < \varepsilon.$$

Then for all large n the probability that either $S^+ > b\sqrt{n}$ or $S_n^- < -b\sqrt{n}$ is less than $2\varepsilon < 1/2$. On the complementary event that $\max(S^{+n}, |S_n^-|) \leq b\sqrt{n}$, we have

$$|Z_n| \leq 8b\sqrt{n}.$$

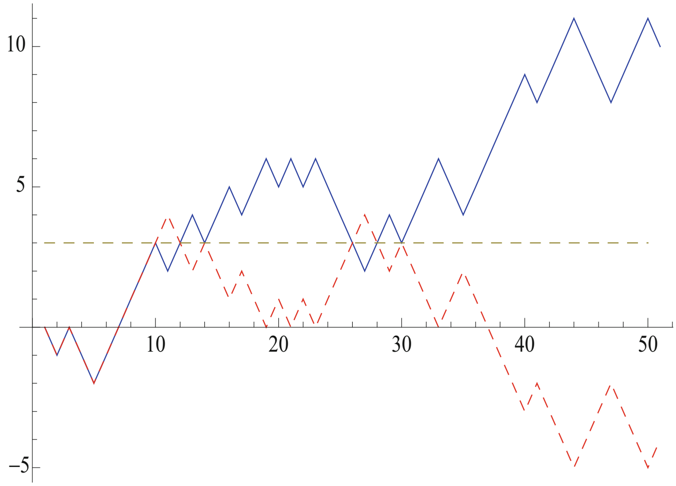


Fig. 1.1 The Reflection Principle

Since this event has probability at least $1 - 2\varepsilon > 1/2$, this implies that $\lim_{n\to\infty} |Z_n|/n$ cannot be positive almost surely. Proposition 1.7.4 follows. □

Proof of Lemma 1.7.5. The proof will rely on the *Reflection Principle* for simple random walk. In its most general form this states that if $(S_n)_{n\geq 0}$ is a simple random walk on $\mathbb{Z}$ with initial point $S_0 = 0$, then so is its *reflection* $(\tilde{S}_n)_{n\geq 0}$ in any horizontal line $y = k$ with $k \in \mathbb{N}$. The reflection $(\tilde{S}_n)_{n\geq 0}$ is defined by

$$\begin{aligned} \tilde{S}_n &= S_n && \text{if } \tau_k \geq n; \\ &= 2k - S_n && \text{if } \tau_k < n \end{aligned} \tag{1.7.6}$$

where

$$\tau_k := \min \{n \in \mathbb{N} \,:\, S_n = k\}\,.$$

See Figure 1.1.

The proof of Lemma 1.7.5 will require only a simple consequence of the Reflection Principle. Let k be a positive integer and r an integer such that $r \leq k$. Let τ_k be as defined above. Then for any $n \in \mathbb{N}$,

$$P\{S_n = r \text{ and } \tau_k \leq n\} = P\{S_n = k + (k - r)\} \tag{1.7.7}$$

Exercise 1.7.6 Prove this.

HINT: Show that there is a one-to-one correspondence between simple random walk paths of length n that terminate at $2k - r$ and paths that visit site k before time n and terminate at r.

The event $\left\{S_n^+ \geq k\right\}$ obviously coincides with the event $\{\tau_k \leq n\}$, so it follows from (1.7.7) (by summing over all integers $r < k$) that

$$P\left\{S_n^+ \geq k\right\} = 2P\{S_n > k\} + P\{S_n = k\}. \tag{1.7.8}$$

Fix $b \in \mathbb{R}_+$, and for any $n \in \mathbb{N}$ let $k = k(b, n)$ be the integer part of $b\sqrt{n}$. The Central Limit Theorem implies that for fixed

$$\lim_{n\to\infty} 2P\{S_n > k\} + P\{S_n = k\} = \frac{2}{\sqrt{2\pi}} \int_b^\infty e^{-t^2/2}\,dt;$$

this proves that the first and third quantities in (1.7.5) are equal. That these equal the second term follows by the symmetry of the simple random walk. □

Exercise 1.7.7 Use the identity (1.7.8) for simple random walk on $\mathbb{Z}$ to prove that $E\tau_1 = \infty$, and use this to show that the expected time of first return to 0 is infinite. HINT: The identity (1.7.8) implies that for any $n \in \mathbb{N}$,

$$P\{\tau_1 > n\} = P\{S_n = 0\} + P\{S_n = 1\}.$$

Now use the fact that for any random variable Y that takes values in the set $\mathbb{Z}_+$ of nonnegative integers,

$$EY = \sum_{n=1}^{\infty} P\{Y \geq n\}.$$

1.8 Excursions of Recurrent Random Walks†

According to Corollary 1.4.6, a recurrent random walk $(X_n)_{n\geq 0}$ almost surely returns to its initial point infinitely many times. Thus, the random variables $T_0, T_1, T_2, \cdots$ defined inductively by

$$\begin{aligned} T_{m+1} &= \min\{n > T_m \,:\, X_n = 1\} && \text{if } N_1 \geq m+1; \\ &= \infty && \text{if } N_1 \leq m \end{aligned} \tag{1.8.1}$$

are almost surely finite. This implies that the trajectory of the random walk can be divided into *excursions* Y_m from the group identity, as follows:

$$\begin{aligned} Y_m &= (X_{T_{m-1}}, X_{T_{m-1}+1}, \cdots, X_{T_m}) && \text{if } T_m < \infty, \\ &= (X_{T_{m-1}}, X_{T_{m-1}+1}, \cdots) && \text{if } T_{m-1} < T_m = \infty, \\ &= \emptyset && \text{if } T_{m-1} = \infty. \end{aligned} \tag{1.8.2}$$

The excursions are random variables that take values in the set $(\cup_{k=0}^{\infty}\Gamma^k) \cup \Gamma^{\infty}$, with $P\{Y_m = \emptyset\} = P\{Y_m \in \Gamma^{\infty}\} = 0$, and by Corollary 1.4.6, the probability that an excursion is infinite is 0.

Lemma 1.8.1 *If the random walk $(X_n)_{n\geq 0}$ is recurrent, then the excursions $Y_1, Y_2, Y_3, \cdots$ are independent and identically distributed.*

Exercise 1.8.2 Prove this.
HINT: What you must show is that for any sequence $y_1, y_2, \cdots$ of elements of the excursion space $\cup_{k=0}^{\infty}\Gamma^k$ and any $m \in \mathbb{N}$,

$$P\left(\bigcap_{i=1}^{m} \{Y_i = y_i\}\right) = \prod_{i=1}^{m} P\{Y_i = y_i\}.$$

It follows (by the *amalgamation principle* — see Proposition A.3.4 in the Appendix) that for any numerical function $f : \cup_{k=0}^{\infty}\Gamma^k \to \mathbb{R}$ the real-valued random variables $f(Y_1), f(Y_2), f(Y_3), \cdots$ are independent and identically distributed. Thus, the fundamental laws governing sums of i.i.d. random variables — the SLLN, the Central Limit Theorem, etc. — apply to sums of the function f over excursions.

Proof of Proposition 1.4.7. Fix a group element x, and define $f(Y_m)$ to be the indicator function of the event that the excursion Y_m visits the state x, that is,

$$f(Y_m) = \mathbf{1}\left\{\sum_{n=T_{m-1}}^{T_m - 1} \mathbf{1}\{X_n = x\} \geq 1\right\}.$$

By Lemma 1.8.1, the random variables $f(Y_m)$ are independent and identically distributed, so the Strong Law of Large Numbers implies that with probability 1,

$$\lim_{m\to\infty} m^{-1} \sum_{k=1}^{m} f(Y_k) = Ef(Y_1) = p,$$

where p is the probability that an excursion makes a visit to x. Consequently, if $p > 0$ then the number $N_x = \sum_{k\geq 1} f(Y_k)$ of visits to x must be $+\infty$ with probability 1. But $p = 0$ is impossible, because this would imply

$$P\{f(Y_m) = 0\} = 1 \quad \text{for every } m \in \mathbb{N} \implies$$
$$P\{f(Y_m) = 0 \text{ for every } m \in \mathbb{N}\} = 1 \implies$$
$$P\{N_x = 0\} = 1,$$

in contradiction to the hypothesis that the random walk is irreducible. □

Exercise 1.8.3 Assume that the random walk $(X_n)_{n\geq 0}$ is not only irreducible and recurrent but also *symmetric*. (A) Show that for any $x \in \Gamma$,

$$E^1 N_x = E^x N_1,$$

where in the second expectation N_1 is the number of visits to 1 before the first return to x. (B) Use this together with the SLLN to show that

$$E^1 N_x = 1.$$

NOTE: This is true for all irreducible, recurrent random walks, not only symmetric walks, but this is harder to prove.

Additional Notes. Polya's Theorem concerning the recurrence/transience dichotomy for random walks on the d-dimensional integer lattices was extended by Chung and Fuchs [25], who gave a Fourier analytic criterion for recurrence of random walks with arbitrary step distributions. Random walks on free groups were first investigated by Dynkin and Maliutov [36], who showed that for nearest-neighbor random walks the *Martin boundary* coicides with the space of ends of the Cayley graph. Random walks on lamplighter groups were introduced by Kaimanovich and Vershik [68]. Woess [133] gives an elegant description of the Cayley graph of the one-dimensional lamplighter lamplighter group $\mathbb{Z}_2 \wr \mathbb{Z}$ and its use in describing the cone of positive harmonic functions for the nearest-neighbor random walk on this group.

Chapter 2
The Ergodic Theorem

2.1 Formulation of the Theorem

If $X_n = \xi_1\xi_2\cdots\xi_n$ is a transient random walk on a finitely generated group Γ with step distribution μ, then the probability q that it will never return to its initial location $X_0 = 1$ is positive. Moreover, for any $n \in \mathbb{N}$ the sequence $(X_n^{-1}X_{n+m})_{m\geq 0}$ is a version of the random walk, as its increments $\xi_{n+1}, \xi_{n+2}, \cdots$ are independent and identically distributed with common distribution μ, and so the probability that it will ever return to the initial state 1 is also $q > 0$. Thus, it is natural to suspect that the *limiting fraction* of times when the random walk leaves its current location never to return is also q. Unfortunately, this cannot be deduced from the Strong Law of Large Numbers (at any rate, not directly), because the indicators of the events

$$F_n := \{\text{no return to } X_n\} := \bigcap_{m>n} \{X_m \neq X_n\} = \bigcap_{m\geq 1} \{\xi_{n+1}\xi_{n+2}\cdots\xi_{n+m} \neq 1\}, \tag{2.1.1}$$

although identically distributed, are not independent. This illustrates a shortcoming of the Strong Law of Large Numbers: it applies to averages of functions $f(\xi_n)$ that depend on a single increment, but not to functions that depend on the entire future trajectory $f(\xi_{n+1}, \xi_{n+2}, \cdots)$. Birkhoff's *Ergodic Theorem* [11] remedies this shortcoming by extending the validity of the Strong Law of Large Numbers to sequences of random variables that are *stationary* and *ergodic*. To define these terms, we must introduce the notion of a *measure-preserving transformation.*

Definition 2.1.1 A *measure-preserving transformation* of a probability space $(\Omega, \mathcal{F}, P)$ is a measurable mapping $T : \Omega \to \Omega$ such that for every event $F \in \mathcal{F}$,

$$P(F) = P(T^{-1}(F)) \quad \iff \quad E\mathbf{1}_F = E(\mathbf{1}_F \circ T). \tag{2.1.2}$$

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_2

An *invariant random variable* for the transformation T is a random variable $Y : \Omega \to \Upsilon$ valued in some measurable space $(\Upsilon, \mathcal{G})$ such that $Y = Y \circ T$; an *invariant event* is an event $F \in \mathcal{F}$ whose indicator function $\mathbf{1}_F$ is an invariant random variable. The set $\mathcal{I}$ of all invariant events F is a σ-algebra, called the *invariant σ-algebra*. The measure-preserving transformation T is *ergodic* if its invariant σ-algebra is *trivial*, in the sense that every event $F \in \mathcal{I}$ has probability either 0 or 1.

Example 2.1.2 For the purposes of the random walker, the most important example of a measure-preserving transformation is the *shift mapping* on an infinite product space. Let $(\Upsilon, \mathcal{G}, \nu)$ be a probability space, and let $Q = \nu \times \nu \times \cdots$ be the corresponding product measure on the sequence space $(\Upsilon^\infty, \mathcal{G}^\infty)$. (Here $\mathcal{G}^\infty$ is the product σ-algebra: see Definition A.3.1 of the Appendix.) The *shift transformation* $\sigma : \Upsilon^\infty \to \Upsilon^\infty$ is the $\mathcal{G}^\infty$-measurable mapping

$$\sigma(x_1, x_2, x_3, \cdots) = (x_2, x_3, x_4, \cdots). \tag{2.1.3}$$

NOTE: The notation $\sigma(Y_1, Y_2, \cdots)$ is also used for the σ-algebra generated by the random variables $Y_1, Y_2, \cdots$, so there is the potential for ambiguity. However, the use of the letter σ is standard for both σ-algebras and the shift mapping; in any case, the meaning will always be clear by context.

Proposition 2.1.3 *The shift mapping $\sigma : \Upsilon^\infty \to \Upsilon^\infty$ is measure-preserving and ergodic for the product measure $Q = \nu \times \nu \times \cdots$.*

Proof. Recall that a *cylinder event* is an element $C \in \mathcal{G}^\infty$ of the form

$$C = G_1 \times G_2 \times \cdots \times G_m \times \Upsilon \times \Upsilon \times \cdots, \tag{2.1.4}$$

where each $G_i \in \mathcal{G}$. The set $\mathcal{A}$ of all finite unions of cylinder events is an *algebra* that generates $\mathcal{G}^\infty$, that is, $\mathcal{G}^\infty$ is the minimal σ-algebra containing $\mathcal{A}$. Consequently, by Corollary A.2.4 of the Appendix, any element of $\mathcal{G}^\infty$ can be arbitrarily well approximated by elements of $\mathcal{A}$: in particular, for any $G \in \mathcal{G}^\infty$ and any $\varepsilon > 0$ there is an event $A \in \mathcal{A}$ such that

$$E_Q|\mathbf{1}_G - \mathbf{1}_A| < \varepsilon.$$

Therefore, to prove that the shift σ is measure-preserving, it suffices to show that for every cylinder event C we have $E_Q \mathbf{1}_C = E_Q \mathbf{1}_C \circ \sigma$. But this follows immediately by the definition of the product measure Q: for any event C of the form (2.1.4),

$$E_Q \mathbf{1}_C = Q(G_1 \times G_2 \times \cdots \times G_m \times \Upsilon \times \Upsilon \times \cdots) = \prod_{i=1}^m \nu(G_i) \quad \text{and}$$

$$E_Q \mathbf{1}_C \circ \sigma = Q(\Upsilon \times G_1 \times G_2 \times \cdots \times G_m \times \Upsilon \times \cdots) = \prod_{i=1}^m \nu(G_i).$$

To prove that the shift σ is ergodic, we will show that every invariant event H satisfies the equality $Q(H) = Q(H)^2$. Fix $\varepsilon > 0$, and let $A \in \mathcal{A}$ be an event whose indicator $\mathbf{1}_A$ is within L^1-distance ε of the indicator $\mathbf{1}_H$, that is, such that $E_Q|\mathbf{1}_A - \mathbf{1}_H| < \varepsilon$. Since A is a finite union of cylinders, there exists $n \in \mathbb{N}$ such that the function $\mathbf{1}_A$ depends only on the first n coordinates of its argument, that is, such that

$$\mathbf{1}_A(x_1, x_2, \cdots, x_n, x_{n+1}, x_{n+2}, \cdots) = \mathbf{1}_A(x_1, x_2, \cdots, x_n, x'_{n+1}, x'_{n+2}, \cdots).$$

Since Q is a product measure, this implies that

$$E_Q \mathbf{1}_A(\mathbf{1}_A \circ \sigma^{n+1}) = Q(A)^2.$$

(Exercise: Check this.) Now if H is an invariant event then

$$E_Q \mathbf{1}_H = E_Q \mathbf{1}_H^2 = E_Q \mathbf{1}_H(\mathbf{1}_H \circ \sigma^{n+1}),$$

so it follows, by the triangle inequality for the L^1-norm and the hypothesis $E_Q|\mathbf{1}_A - \mathbf{1}_H| < \varepsilon$, that

$$|E_Q \mathbf{1}_H - E_Q \mathbf{1}_H^2| \leq |E_Q \mathbf{1}_A - E_Q \mathbf{1}_A(\mathbf{1}_A \circ \sigma^{n+1})| + 3\varepsilon = 3\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this proves that $Q(H) = Q(H)^2$. □

Example 2.1.4 Sometimes, interesting functionals of random walks involve two-sided sequences, so the shift on two-sided sequence space is also of some importance. As in the preceding example, let $(\Upsilon, \mathcal{G}, \nu)$ be a probability space, and let Q be the corresponding product measure on the two-sided sequence space $(\Upsilon^{\mathbb{Z}}, \mathcal{G}^{\mathbb{Z}})$. The shift mapping $\sigma : \Upsilon^{\mathbb{Z}} \to \Upsilon^{\mathbb{Z}}$ is defined in the obvious way, and once again is an ergodic, measure-preserving transformation.

If T is a measure-preserving transformation of a probability space $(\Omega, \mathcal{F}, P)$ then by definition the equality $E\mathbf{1}_F = E(\mathbf{1}_F \circ T)$ holds for all events $F \in \mathcal{F}$. This equality extends to linear combinations of indicators, by the linearity of expectation, and hence, by the Monotone Convergence Theorem, to all nonnegative random variables. Therefore, for every integrable real-valued random variable Y,

$$EY = E(Y \circ T). \tag{2.1.5}$$

This, in turn, implies that the mapping $Y \mapsto Y \circ T$ is a *linear isometry* of every L^p space, that is, it preserves the L^p-norm $\|\cdot\|_p$.

Definition 2.1.5 Let $\mathbf{Y} = (Y_n)_{n \in \mathbb{N}}$ (or $\mathbf{Y} = (Y_n)_{n \in \mathbb{Z}}$) be a sequence of random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$ that take values in a measurable space $(\Upsilon, \mathcal{G})$. The *distribution* (sometimes called the *joint distribution* or *law*) of the sequence is the induced probability measure $Q := P \circ \mathbf{Y}^{-1}$ on

the product space $(\Upsilon^{\mathbb{N}}, \mathcal{G}^{\mathbb{N}})$ (or $(\Upsilon^{\mathbb{Z}}, \mathcal{G}^{\mathbb{Z}})$, if the sequence $\mathbf{Y}$ is two-sided). The sequence $\mathbf{Y}$ is said to be *stationary* if the shift mapping σ on the relevant sequence space preserves the joint distribution Q, and it is *ergodic* if the shift is ergodic for Q.

It follows from (2.1.5) that if the sequence $\mathbf{Y} = (Y_n)_{n \in \mathbb{N}}$ is stationary, then for any random variable $f : \Upsilon \to \mathbb{R}$ such that $E|f(Y_1)| < \infty$,

$$Ef(Y_1) = Ef(Y_2) = Ef(Y_3) = \cdots . \tag{2.1.6}$$

Theorem 2.1.6 (Birkhoff's Ergodic Theorem) *If T is an ergodic, measure-preserving transformation of a probability space $(\Omega, \mathcal{F}, P)$, then for every real-valued random variable $Z : \Omega \to \mathbb{R}$ with finite first moment $E|Z| < \infty$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} Z \circ T^i = EZ \quad \textit{almost surely.} \tag{2.1.7}$$

Equivalently, if $\mathbf{Y} = (Y_n)_{n \in \mathbb{N}}$ is an ergodic, stationary sequence of random variables valued in a measurable space $(\Upsilon, \mathcal{G})$, then for every product-measurable $f : \Upsilon^{\mathbb{N}} \to \mathbb{R}$ such that $E|f(\mathbf{Y})| < \infty$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ \sigma^i(\mathbf{Y}) = Ef(\mathbf{Y}). \tag{2.1.8}$$

We defer the proof to Section 2.4, after we have explored several of the theorem's consequences. Birkhoff's theorem applies, of course, when the sequence $(Y_n)_{n\geq 1}$ consists of independent, identically distributed random variables, since any such sequence is stationary and ergodic, by Example 2.1.2. Therefore, the Strong Law of Large Numbers (Theorem 1.6.3) is a special case of Theorem 2.1.6.

Exercise 2.1.7 Show that under the hypotheses of Theorem 2.1.6, for every real number $C < \infty$, with probability one,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=m+1}^{m+n} Z \circ T^i = EZ \tag{2.1.9}$$

uniformly for $0 \leq m \leq Cn$, that is,

$$\lim_{n\to\infty} \max_{0\leq m\leq Cn} \left| \frac{1}{n} \sum_{i=m+1}^{m+n} Z \circ T^i - EZ \right| = 0 \quad \text{almost surely.} \tag{2.1.10}$$

2.2 The Range of a Random Walk

Theorem 2.1.6 applies to the indicators of the sequence F_n of events defined by (2.1.1), as these are of the form $f \circ \sigma^n(\boldsymbol{\xi})$, where $\boldsymbol{\xi} = (\xi_n)_{n \geq 1}$ is the sequence of increments of the random walk X_n. Let's state the result formally as follows.

Corollary 2.2.1 *Let $(X_n)_{n \geq 0}$ be a random walk (not necessarily transient) on a finitely generated group Γ, and for each $n \geq 0$ let F_n be the event that the random walk never returns to the site X_n, that is, $F_n = \cap_{m>n} \{X_m \neq X_n\}$. Then with probability one,*

$$\lim_{n \to \infty} n^{-1} \sum_{j=0}^{n} \mathbf{1}_{F_j} = P(F_0). \tag{2.2.1}$$

Kesten, Spitzer, and Whitman (cf. [117], Section I.4) and later, in more generality, Derriennic [32], observed that this has interesting ramifications for the *range* of the random walk $(X_n)_{n \geq 0}$. Define, for each nonnegative integer n, the random variable R_n to be the number of distinct sites visited by time n, formally,

$$R_n := |\{X_0, X_1, \cdots, X_n\}| = |\{1, \xi_1, \xi_1\xi_2, \cdots, \xi_1\xi_2 \cdots \xi_n\}|. \tag{2.2.2}$$

Theorem 2.2.2 *For any random walk $(X_n)_{n \geq 0}$ on any finitely generated group Γ,*

$$\lim_{n \to \infty} \frac{R_n}{n} = P\{\textit{no return to } 1\} := P(F_0) \quad \textit{almost surely.} \tag{2.2.3}$$

By definition, a random walk X_n on a finitely generated group is *recurrent* if and only if the probability of return to the initial point is 1, equivalently, if $P\{\text{no return to } 1\} = 0$. Consequently, the Kesten-Spitzer-Whitman theorem implies that X_n is recurrent if and only if the range R_n grows *sublinearly* with n.

Corollary 2.2.3 *Any random walk $S_n = \sum_{i=1}^{n} \xi_i$ on the group $\mathbb{Z}$ whose step distribution has finite first moment $E|\xi_1|$ and mean $E\xi_1 = 0$ is recurrent.*

Proof of Corollary 2.2.3 If $E\xi_i = 0$, then the Strong Law of Large Numbers implies that the sequence S_n/n converges to 0 almost surely. This implies that for any $\varepsilon > 0$, with probability 1 all but at most finitely many terms of the sequence S_n/n fall in the interval $(-\varepsilon, \varepsilon)$; and this in turn implies that for large n the range R_n will satisfy

$$R_n \leq 2n\varepsilon.$$

Since $\varepsilon >$ is arbitrary, it follows that $R_n/n \to 0$ almost surely. Consequently, by Theorem 2.2.2, the probability of no return is 0, and so the random walk must be recurrent. □

Proof of Theorem 2.2.2 The random variable R_n is the number of distinct sites visited by time n; these are in one-to-one correspondence with the times $j \le n$ that the random walk is at a site X_j that will not be visited again by time n. Hence,

$$\begin{aligned} R_n &= \sum_{j=0}^{n} \mathbf{1}\left\{X_j \text{ not revisited by time } n\right\} \\ &\ge \sum_{j=0}^{n} \mathbf{1}\left\{X_j \text{ never revisited}\right\} = \sum_{j=0}^{n} \mathbf{1}_{F_j}, \end{aligned}$$

and so it follows by Corollary 2.2.1 that with probability one,

$$\liminf_{n\to\infty} \frac{R_n}{n} \ge P(F_0).$$

Since the random variables $\xi_1, \xi_2, \cdots$ are i.i.d., the events in the last sum all have the same probability, specifically, the probability that the random walk never returns to its initial site 1. Thus, taking expectations and letting $n \to \infty$ gives

$$\liminf_{n\to\infty} ER_n/n \ge P\{\text{no return to } 1\}.$$

A similar argument proves the upper bound. Fix an integer $k \ge 1$; then for any $0 \le j \le n-k$, the event that X_j is not revisited by time n is contained in the event that X_j is not revisited in the k steps following time j. Consequently,

$$\begin{aligned} R_n &= \sum_{j=0}^{n} \mathbf{1}\left\{X_j \text{ not revisited by time } n\right\} \\ &\le \sum_{j=0}^{n-k} \mathbf{1}\left\{X_j \text{ not revisited in next } k \text{ steps}\right\} + k \\ &= \sum_{j=0}^{n-k} \mathbf{1}\left\{\xi_{j+1}\xi_{j+2}\cdots\xi_{j+l} \ne 1 \ \forall\, l \le k\right\} + k \end{aligned}$$

The Ergodic Theorem applies to the last sum in this chain, since the sequence $\boldsymbol{\xi} = (\xi_n)_{n\in\mathbb{N}}$ is stationary and ergodic: thus, with probability one,

$$\lim_{n\to\infty} n^{-1} \sum_{j=0}^{n-k} \mathbf{1}\left\{\xi_{j+1}\xi_{j+2}\cdots\xi_{j+l} \ne 1 \ \forall\, l \le k\right\} = P\left(\bigcap_{l=1}^{k} \{X_l \ne 1\}\right).$$

It now follows that for any $k \ge 1$, with probability one,

$$\limsup_{n\to\infty} \frac{R_n}{n} \leq P\left(\bigcap_{l=1}^{k} \{X_l \neq 1\}\right).$$

But as $k \to \infty$, these probabilities decrease to the probability of no return, by the dominated convergence theorem, that is,

$$\lim_{k\to\infty} P\left(\bigcap_{l=1}^{k} \{X_l \neq 1\}\right) = P(F_0).$$

Hence, with probability one, the liminf and limsup are both equal to $P(F_0)$. □

Remark 2.2.4 Nowhere in this section was the full strength of the hypothesis that the increments ξ_n are i.i.d. used: all that was really needed is that the sequence $(\xi_n)_{n\geq 1}$ is stationary and ergodic, since that is all the Ergodic Theorem requires. Therefore, the conclusions of both Theorem 2.2.2 and Corollary 2.2.3 are valid under the weaker hypothesis that the increments ξ_n form a stationary, ergodic sequence.

Exercise 2.2.5 Let $(Y_n)_{n\in\mathbb{N}}$ be a stationary, ergodic sequence of real-valued random variables such that $E|Y_1| < \infty$, and let $S_n = \sum_{i=1}^{n} Y_i$.

(A) Show that the event $\{\lim_{n\to\infty} S_n = +\infty\}$ has probability 0 or 1.
(B) Show that if $P\{\lim_{n\to\infty} S_n = +\infty\} = 1$ then $EY_1 > 0$.

HINT: For part (B), show that if $P\{\lim_{n\to\infty} S_n = +\infty\} = 1$ then for some $\varepsilon > 0$ the event $\left\{\min_{n\geq 1} S_n \geq \varepsilon\right\}$ must have positive probability, and use this to show that $P\left\{\liminf_{n\geq 1} S_n/n > 0\right\} = 1$.

2.3 Cut Points of a Random Walk

If a random walk is transient, then each point along its trajectory must be visited for a last time. This, however, does not preclude the possibility that for each time $n \in \mathbb{N}$ there will be a return to one of the R_n points in the set $\{X_i\}_{i\leq n}$ at some future time $n+m$. Are there random walks for which there are times beyond which the random walk no longer returns to any of its previously visited sites?

Definition 2.3.1 Let X_n be a transient random walk on a finitely generated group Γ. For any integer $n \geq 0$, say that the pair (n, X_n) is a *cut point* of the random walk if

$$\{X_0, X_1, X_2, \cdots, X_n\} \cap \{X_{n+1}, X_{n+2}, \cdots\} = \emptyset. \tag{2.3.1}$$

Define the *cut point event* K_n to be the event that (n, X_n) is a cut point.

Theorem 2.3.2 *Let $(X_n)_{n\geq 0}$ be a transient random walk on Γ with step distribution μ and initial point $X_0 = 1$, and let $(Y_n)_{n\geq 0}$ be an independent random walk with step distribution $\hat{\mu}$ and initial point $Y_0 = 1$, where $\hat{\mu}$ is the reflection of μ. If K_n is the cut point event* (2.3.1) *for the random walk $(X_n)_{n\geq 0}$, then*

$$\lim_{m\to\infty} P(K_m) = \alpha \quad \textit{and} \tag{2.3.2}$$

$$\lim_{n\to\infty} n^{-1} \sum_{m=1}^{n} \mathbf{1}_{K_m} = \alpha \quad \textit{almost surely,} \tag{2.3.3}$$

where α is the probability that the sets $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 0}$ do not intersect.

Recall (cf. Exercise 1.4.11) that the *reflection* $\hat{\mu}$ of a probability distribution μ on Γ is the unique probability distribution such that $\hat{\mu}(g) = \mu(g^{-1})$ for every $g \in \Gamma$. Thus, if $\zeta_1, \zeta_2, \cdots$ are independent, identically distributed random variables with distribution μ then the sequence

$$Y_n := \zeta_1^{-1}\zeta_2^{-1}\cdots\zeta_n^{-1}$$

is a random walk with step distribution $\hat{\mu}$. It follows that independent random walks $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ with step distributions μ and $\hat{\mu}$, respectively, can be defined on any probability space that supports a doubly-infinite sequence of random variables with distribution μ. Consequently, there is no loss of generality in assuming that the ambient probability space is the two-sided product space $\Gamma^{\mathbb{Z}}$.

Assumption 2.3.3 *Assume that the probability space $(\Omega, \mathcal{F}, P)$ is the two-sided sequence space $\Omega = \Gamma^{\mathbb{Z}}$ with σ-algebra $\mathcal{F}$ generated by the cylinder sets and $P = P_\mu$ the product measure. Denote by $\xi_n : \Omega \to \Gamma$ the coordinate random variables, and for each $n \in \mathbb{Z}$ define*

$$\begin{aligned} X_n &= \xi_1\xi_2\cdots\xi_n && \textit{if } n \geq 0; \\ X_n &= \xi_0^{-1}\xi_{-1}^{-1}\xi_{-2}^{-1}\cdots\xi_{n+1}^{-1} && \textit{if } n \leq -1; \quad \textit{and} \\ Y_n &= X_{-n} && \textit{for } n \geq 0. \end{aligned} \tag{2.3.4}$$

Proof of Theorem 2.3.2 Let X_n, Y_n, ξ_n be as in Assumption 2.3.3, and let $\sigma : \Gamma^{\mathbb{Z}} \to \Gamma^{\mathbb{Z}}$ be the forward shift operator. For each $m \in \mathbb{N}$, define F_m to be the event that the forward trajectory $\{X_n\}_{n\geq 1}$ does not intersect the set $\{Y_n\}_{0\leq n\leq m}$, and let $F_\infty = \cap_{m\geq 1} F_m$. The events F_m are decreasing in m, so

$$\alpha := P(F_\infty) = \lim_{m\to\infty} \downarrow P(F_m). \tag{2.3.5}$$

For any $m \in \mathbb{N}$,

$$\{X_0, X_1, X_2, \cdots, X_m\} \cap \{X_{m+1}, X_{m+2}, \cdots\} = \emptyset \quad \Longleftrightarrow$$

$$\left\{X_m^{-1}X_0, X_m^{-1}X_1, X_m^{-1}X_2, \cdots, 1\right\} \cap \left\{X_m^{-1}X_{m+1}, X_m^{-1}X_{m+2}, \cdots\right\} = \emptyset,$$

and so $\mathbf{1}_{K_m} = \mathbf{1}_{F_m} \circ \sigma^m$. Therefore, $P(K_m) = P(F_m)$, whence (2.3.2) follows from (2.3.5).

Since the events F_m are decreasing in m, we have, for any positive integers $m \leq k$,

$$\sum_{n=1}^{k} \mathbf{1}_{F_\infty} \circ \sigma^n \leq \sum_{n=1}^{k} \mathbf{1}_{K_n} \leq m + \sum_{n=m+1}^{k} \mathbf{1}_{F_m} \circ \sigma^n. \tag{2.3.6}$$

By the Ergodic Theorem,

$$\lim_{k\to\infty} k^{-1} \sum_{i=0}^{k} \mathbf{1}_{F_m} \circ \sigma^i = P(F_m) \quad \text{and}$$

$$\lim_{k\to\infty} k^{-1} \sum_{i=0}^{k} \mathbf{1}_{F_\infty} \circ \sigma^i = P(F_\infty) \quad \text{almost surely.}$$

Consequently, since $P(F_m) \downarrow P(F_\infty)$, the relation (2.3.3) follows from the inequalities (2.3.6). □

Exercise 2.3.4 Show that for any transient, symmetric random walk X_n on a finitely generated group Γ, the limiting fraction of integers n such that the site X_n is visited only once is almost surely equal to $(P\,\{\text{no return to } 1\})^2$.

Exercise 2.3.5

(A) Show that for any transient, irreducible random walk $(X_n)_{n\geq 0}$ on a finitely generated group Γ there is a sequence of positive scalars α_K such that with probability one,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n} |\mathbb{B}_K(X_m) \cap \{1, X_1, X_2, \cdots, X_m\}| = \alpha_K.$$

(Recall that $\mathbb{B}_K(g)$ is the ball of radius K centered at g in the Cayley graph.)

(B) Show that if the random walk is also symmetric, then

$$\alpha_K = \sum_{n\geq 0} \sum_{x\in\mathbb{B}_K(1)} \mu^{*n}(x).$$

The next obvious question is this: for which transient random walks is $\alpha > 0$? In general, the question is somewhat delicate, but there is one large class of random walks for which we can prove without great difficulty that $\alpha > 0$.

Definition 2.3.6 The *spectral radius* of a random walk X_n on a finitely generated group Γ is defined by

$$\varrho := \limsup_{n\to\infty} P\{X_{kn} = 1\}^{1/kn}, \tag{2.3.7}$$

where $k \in \mathbb{N}$ is the *period* of the random walk, that is, k is the greatest common divisor of the set of all $n \in \mathbb{Z}_+$ such that $P\{X_n = 1\} > 0$.

If the period of the random walk is k, then $P\{X_n = 1\} = 0$ unless n is a multiple of k, so we could just as well have defined ϱ to be the limsup of the sequence $P\{X_n = 1\}^{1/n}$. However, in Section 3.2, we (actually, you) will show that the limsup in (2.3.7) (that is, along the sequence nk) is really a limit. In Chapter 4 (cf. Proposition 4.4.10), we will show that ϱ is the norm of the *Markov operator* associated to the random walk: this explains the use of the term *spectral radius*. In Chapter 5 we will show that for every irreducible, symmetric random walk on a *nonamenable*, finitely generated group (for instance, the free group $\mathbb{F}_d$ for any $d \geq 2$, or the matrix group $SL(n, \mathbb{Z})$, for any $n \geq 2$), the spectral radius is less than 1.

Proposition 2.3.7 *Assume that the step distribution μ is symmetric and has support either $\mathbb{A}$ or $\mathbb{A} \cup \{1\}$, for some finite, symmetric generating set $\mathbb{A}$. If $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ are independent random walks with step distribution μ and initial points $X_0 = Y_0 = 1$, and if the spectral radius is less than 1 then*

$$P(\{X_n\}_{n\geq 1} \cap \{Y_n\}_{n\geq 0} = \emptyset) > 0. \tag{2.3.8}$$

The proof will rely on the following auxiliary estimate.

Lemma 2.3.8 *Under the hypotheses of Proposition 2.3.7, there exist constants $0 < r < 1$ and $C < \infty$ such that for every $x \in \Gamma$,*

$$P(\{X_n\}_{n\geq 0} \cap \{xY_n\}_{n\geq 0} \neq \emptyset) \leq Cr^{|x|}. \tag{2.3.9}$$

Proof. By hypothesis, the spectral radius is less than one, so there exists $r \in [\varrho, 1)$ such that $P\{X_n = 1\} \leq r^n$ for every $n \geq 0$. Since the random walk is symmetric, for any $x \in \Gamma$ and any $n \geq 0$, the probability that the random walk reaches x in n steps is the same as the (conditional) probability that the next n steps take it back to 1; consequently,

$$P\{X_n = x\}^2 \leq P\{X_{2n} = 1\} \leq r^{2n}.$$

By assumption, the random walks $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ are independent. Hence, the symmetry of the step distribution implies that for any integers $m, n \geq 0$ the random variable $Y_m^{-1}X_n$ has the same distribution as does X_{m+n}. It follows that for any $x \in \Gamma$,

$$P\{X_n = xY_m\} = P\left\{Y_m^{-1}X_n = x\right\} = P\{X_{m+n} = x\} \le r^{m+n}.$$

Now since the random walk $(X_n)_{n\ge0}$ is nearest-neighbor, the site x cannot be reached in fewer than $|x|$ steps, so $P\{X_{m+n} = x\} = 0$ unless $m + n \ge |x|$. Therefore, for a suitable constant $C < \infty$,

$$P\{X_n = xY_m \text{ for some } n, m \ge 0\} \le \sum_{m,n\,:\,m+n\ge|x|} r^{m+n} \le Cr^{|x|/2}.$$

□

Proof of Proposition 2.3.7 Assume that the sequences $(X_n)_{n\in\mathbb{Z}}$ and $(Y_n)_{n\ge0}$ are defined by (2.3.4), where the random variables $(\xi_n)_{n\in\mathbb{Z}}$ are independent and identically distributed, with common distribution μ. Fix $m \in \mathbb{N}$ so large that

$$Cr^{2m} + 2Cr^m/(1-r) < 1, \tag{2.3.10}$$

where C, r are constants in the inequalities (2.3.9). Let $\gamma = (x_{-m}, x_{-m+1}, \cdots, x_m)$ be a *geodesic path* in Γ of length $2m + 1$ such that $x_0 = 1$. (Recall that a geodesic path in Γ is a sequence of group elements x_i such that for any two points x_i, x_j in the sequence, $d(x_i, x_j) = |i - j|$. For the existence of doubly-infinite geodesic paths with $x_0 = 1$, see Exercise 1.2.5.) Define G to be the event that the segment $(X_n)_{|n|\le m}$ of the random walk path follows the geodesic path γ, that is,

$$\begin{aligned} G &:= \{X_n = x_n \text{ for all } |n| \le m\} \\ &= \{X_n = x_n \text{ for all } 0 \le n \le m\} \cap \{Y_n = x_{-n} \text{ for all } 0 \le n \le m\}. \end{aligned}$$

This event depends only on the increments $(\xi_n)_{-m<n\le m}$, and because the support of the step distribution contains the generating set $\mathbb{A}$, the event has positive probability $P(G)$.

For each integer $n \ge 0$ define

$$\begin{aligned} X_n^* &= X_m^{-1}X_{m+n} = \xi_{m+1}\xi_{m+2}\cdots\xi_{m+n} \quad \text{and} \\ Y_n^* &= Y_m^{-1}Y_{m+n} = \xi_{-m}\xi_{-m-1}\cdots\xi_{-m-n+1}. \end{aligned}$$

The sequences $(X_n^*)_{n\ge0}$ and $(Y_n^*)_{n\ge0}$ are independent versions of the random walk that are mutually independent of the increments $(\xi_n)_{-m<n\le m}$, and hence independent of the event G; furthermore, on the event G,

$$\begin{aligned} X_n &= x_m X_n^* \quad \text{and} \\ Y_n &= x_{-m} Y_n^* \quad \text{for all } n \ge 0. \end{aligned}$$

Consequently, the events

$$F_1 = \left\{ \left\{ x_m X_n^* \right\}_{n \geq 0} \cap \left\{ x_{-j} \right\}_{0 \leq j \leq m} \neq \emptyset \right\},$$

$$F_2 = \left\{ \left\{ x_{-m} Y_n^* \right\}_{n \geq 0} \cap \left\{ x_j \right\}_{0 \leq j \leq m} \neq \emptyset \right\}, \quad \text{and}$$

$$F_3 = \left\{ \left\{ x_m X_n^* \right\}_{n \geq 0} \cap \left\{ x_{-m} Y_n^* \right\}_{n \geq 0} \neq \emptyset \right\},$$

are independent of G, and on the event $G \cap (F_1 \cup F_2 \cup F_3)^c$ the trajectories $\{X_n\}_{n \geq 1}$ and $\{Y_n\}_{n \geq 0}$ do not overlap. Therefore, to complete the proof it suffices to show that

$$P(F_1) + P(F_2) + P(F_3) < 1.$$

Since γ is a geodesic path, for any $0 \leq j \leq m$ the distances $d(x_{-j}, x_m)$ and $d(x_j, x_{-m})$ both equal $m + j$. Hence, by Lemma 2.3.8, the probability that the random walk $(x_m X_n^*)_{n \geq 0}$ (respectively, the random walk $(x_{-m} Y_n^*)_{n \geq 0}$) ever visits x_{-j} (respectively, x_j) is no larger than Cr^{m+j}. It follows that

$$P(F_i) \leq C \sum_{j=0}^{\infty} r^{m+j} = Cr^m/(1-r) \quad \text{for each } \; i = 1, 2.$$

Similarly, because the distance between x_{-m} and x_m is $2m$,

$$P(F_3) < Cr^{2m.}$$

Thus, $P(F_1) + P(F_2) + P(F_3) < 1$, by (2.3.10). □

Exercise 2.3.9 Use the Local Central Limit Theorem and Hoeffding's inequality to show that if $d \in \mathbb{N}$ is sufficiently large then the probability α that two independent simple random walks on the integer lattice $\mathbb{Z}^d$ will not intersect is positive. Consequently, simple random walk in large dimensions has cut points.
NOTE: The Local CLT implies that for a simple random walk $(X_n)_{n \geq 0}$ in $\mathbb{Z}^d$

$$P\{S_{2n} = x\} \sim \exp\left\{-|x|^2/2n\right\}/(4\pi dn)^{d/2} \tag{2.3.11}$$

uniformly for all $x \in H$ such that $|x|^2/n \leq C$, where H is the index-2 subgroup of $\mathbb{Z}^d$ containing all points x of even parity and $C < \infty$ is any finite constant.

Example 2.3.10 (Nearest-Neighbor Random Walks on Trees) Let's show that any transient, nearest-neighbor random walk on a group whose Cayley graph is an infinite, regular tree has cut points. These occur whenever the random walk visits a new point x and subsequently fails to re-cross the edge incident to x on the unique self-avoiding path from 1 to x.

For ease of exposition, consider the special case $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$, the free product of three copies of the two-element group $\mathbb{Z}_2$. Elements of Γ are reduced words in the letters a, b, c (that is, finite sequences of letters in which no letter appears twice

in succession). Let $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be independent, nearest-neighbor random walks on Γ with common step distribution $P\{X_1 = i\} = P\{Y_1 = i\} = \mu(i)$, where $i \in \{a, b, c\}$. Assume that the random walks are irreducible, that is, $\mu(i) > 0$ for each $i = a, b, c$.

Denote by $h : \Gamma \to [0, 1]$ the *hitting probability function* of the random walk, defined by

$$\begin{aligned} h(x) &= P^1\{X_n = x \text{ for some } n \geq 0\} \\ &= P^x\{X_n = 1 \text{ for some } n \geq 0\}. \end{aligned} \tag{2.3.12}$$

Clearly, $h(i) \geq \mu(i) > 0$ for each $i = a, b, c$ (cf. Exercise 1.6.5(B)). Furthermore, $h(i) < 1$, because otherwise, by (2.3.12) and the Markov Property, the random walk would cross and re-cross the edge of the Cayley graph connecting 1 and i infinitely often, with probability one; this is impossible, because any irreducible random walk on the group $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$ is transient.

For each of the generators $i \in \{a, b, c\}$ the probability that the random walk $(X_n)_{n\geq 0}$ (or the random walk $(Y_n)_{n\geq 0}$) makes its first step accross the edge from 1 to i and then never re-crosses this edge is $\mu(i)(1 - h(i)) > 0$. Consequently, since the random walks $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ are independent, the probability α that their paths intersect only at the initial point $X_0 = Y_0 = 1$ satisfies

$$\alpha \geq \sum_{i \neq j} \mu(i)\mu(j)(1 - h(i))(1 - h(j)) > 0.$$

Therefore, by Theorem 2.3.2, the limiting frequency (2.3.3) of cut points in $(X_n)_{n\geq 0}$ is positive.

2.4 Proof of the Ergodic Theorem

Like many almost everwhere convergence theorems in probability and analysis, the Ergodic Theorem relies on a *maximal inequality*, and this inequality, in turn, relies on a *covering lemma*. Here the covering lemma takes an especially simple form. This is because finite intervals in the set of nonnegative integers have well-defined minima: in particular, every finite interval $J \subset \mathbb{Z}$ has the form $J = [a, b] := \{n \in \mathbb{Z} : a \leq n \leq b\}$.

Lemma 2.4.1 (Covering Lemma) *Let $\mathcal{J}$ be a collection of finite intervals of nonnegative integers, and let $L = \{a \in \mathbb{Z}_+ : J = [a, b]$ for some $J \in \mathcal{J}\}$ be the set of left endpoints of intervals in $\mathcal{J}$. Then there is a subset $\mathcal{J}_* \subset \mathcal{J}$ consisting of nonoverlapping intervals such that*

$$L \subset \bigcup_{J \in \mathcal{J}_*} J. \tag{2.4.1}$$

Proof. Let $a_0 = \min L$ be the smallest element of L; since $L \subset \mathbb{Z}_+$, this is well-defined. Choose any interval $J_0 = [a_0, b_0]$ with a_0 as its left endpoint. Now define intervals $J_1, J_2, \cdots$ inductively so that for each $m \geq 0$, the left endpoint a_m of J_m satisfies

$$a_m = \min(L \setminus \cup_{i=0}^{m-1} J_i),$$

that is, so that a_m is the smallest element of L not contained in any of the intervals J_i with index $i \leq m-1$ (unless $L \subset \cup_{i=0}^{m-1} J_i$, in which case the induction ends). The collection $\mathcal{J}_* := \{J_i\}_{i \geq 0}$ obviously has the desired properties. □

Proposition 2.4.2 (Wiener's Maximal Inequality) *If T is a measure-preserving transformation of the probability space $(\Omega, \mathcal{F}, P)$ then for every nonnegative real-valued random variable Z and every real number $\alpha > 0$,*

$$P\left\{\sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=0}^{n-1} Z \circ T^i > \alpha\right\} \leq \frac{EZ}{\alpha}. \tag{2.4.2}$$

Proof. Fix an integer $K \geq 1$, and define the event

$$F = F_{K,\alpha} = \left\{\max_{k \leq K} \frac{1}{k} \sum_{i=0}^{k-1} Z \circ T^i > \alpha\right\}.$$

For any $\omega \in \Omega$, define $L = L(\omega) \subset \mathbb{Z}_+$ to be the set of all integers $n \geq 0$ such that $T^n(\omega) \in F$. By definition, if $n \in L(\omega)$ then for some integer k satisfying $1 \leq k \leq K$,

$$\sum_{i=0}^{k-1} Z \circ T^{n+i}(\omega) > k\alpha. \tag{2.4.3}$$

Now define $\mathcal{J} = \mathcal{J}(\omega)$ to be the collection of all intervals $[n, n+k)$ with $1 \leq k \leq K$ such that (2.4.3) holds. Clearly, L coincides with the set of left endpoints of intervals in $\mathcal{J}$; thus, by Lemma 2.4.1, there is a subset $\mathcal{J}_* = \mathcal{J}_*(\omega)$ of $\mathcal{J}$ consisting of pairwise disjoint intervals of lengths $\leq K$ such that for each interval $J \in \mathcal{J}_*(\omega)$,

$$\sum_{i \in J} Z \circ T^i(\omega) > \alpha|J|, \tag{2.4.4}$$

where $|J|$ denotes the cardinality of J.

By construction, every integer $n \in L(\omega)$ is contained in one of the intervals in the collection $\mathcal{J}_*$. Since each of these intervals has length $\leq K$, and since Z is nonnegative, it follows by (2.4.4) that for every $n \geq 1$,

$$\begin{aligned}
\sum_{i=0}^{nK} Z \circ T^i(\omega) &\geq \sum_{J \in \mathcal{J}_*(\omega) \text{ and } J \subset [0,nK]} \sum_{i \in J} Z \circ T^i(\omega) \\
&\geq \sum_{J \in \mathcal{J}_*(\omega) \text{ and } J \subset [0,nK]} \alpha |J| \\
&\geq \alpha \sum_{i=0}^{nK-K-1} \mathbf{1}_F \circ T^i(\omega).
\end{aligned}$$

Taking expectations and using the hypothesis that T is measure-preserving, we find that

$$(nK)EZ \geq (nK - K)\alpha P(F).$$

Dividing by nK and sending $n \to \infty$, we conclude that

$$P(F) = P(F_{K,\alpha}) \leq EZ/\alpha.$$

Since K is arbitrary, an application of the monotone convergence theorem yields the inequality (2.4.2). $\square$

Proof of the Ergodic Theorem for bounded Z. If Z is bounded, then without loss of generality we can assume that $0 \leq Z \leq 1$, because rescaling and addition or subtraction of a constant from Z does not affect the validity of the theorem. Set

$$\begin{aligned}
A_n &= \frac{1}{n} \sum_{i=0}^{n-1} Z \circ T^i, \\
A^* &= \limsup_{n \to \infty} A_n, \quad \text{and} \\
A_* &= \liminf_{n \to \infty} A_n.
\end{aligned}$$

The random variables A^* and A_* are invariant, since the random variable Z is bounded, and consequently, since the measure-preserving transformation T is ergodic, A^* and A_* are almost surely constant, with values $a_* \leq a^*$. By Wiener's maximal inequality, for any $\alpha > 0$,

$$P\left\{A^* > \alpha\right\} \leq EZ/\alpha.$$

But since A^* is almost surely constant with value a^*, the probability on the left must be either 0 or 1, and so for any $\alpha > EZ$ the probability that $A^* > \alpha$ is 0. Therefore, $a^* \leq EZ$. To obtain the matching inequality $a_* \geq EZ$, replace Z by $1 - Z$. $\square$

Proof of the Ergodic Theorem for bounded Z. It suffices to consider the case where Z is nonnegative, as the general case can then be deduced by decomposition into positive and negative parts. Without loss of generality, assume that $EZ = 1$, and let $\beta > 1$ be a (large) constant. Since the Ergodic Theorem holds for bounded random variables, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (Z \wedge \beta) \circ T^i = E(Z \wedge \beta) \quad \text{almost surely.} \tag{2.4.5}$$

On the other hand, by Wiener's maximal inequality, if $Y = Z\mathbf{1}\{Z \geq \beta\}$ then

$$P\left\{\sup_{n\geq 1} n^{-1} \sum_{i=0}^{n-1} Y \circ T^i \geq \alpha\right\} \leq \frac{EZ\mathbf{1}\{Z \geq \beta\}}{\alpha}.$$

By the dominated convergence theorem, for any $\alpha > 0$, no matter how small, there exists $\beta < \infty$ large enough that the right side of this inequality is less than α, that is,

$$EZ\mathbf{1}\{Z \geq \beta\} < \alpha^2.$$

Hence, by (2.4.5),

$$P\left\{\sup_{n\geq 1} n^{-1} \sum_{i=0}^{n-1} (Z - Z \wedge \beta) \circ T^i \geq \alpha\right\} \leq \alpha \quad \Longrightarrow$$

$$P\left\{\limsup_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} Z \circ T^i \geq E(Z \wedge \beta) + \alpha\right\} \leq \alpha \quad \text{and}$$

$$P\left\{\liminf_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} Z \circ T^i \leq E(Z \wedge \beta) + \alpha\right\} = 0.$$

Since $\alpha > 0$ can be chosen arbitrarily small, it follows that the limit

$$\lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} Z \circ T^i$$

exists almost surely and equals $\lim_{\beta\to\infty} E(Z \wedge \beta) = EZ$. $\square$

2.5 Non-Ergodic Transformations†

Theorem 2.1.6 requires that the underlying measure-preserving transformation be *ergodic*. What, if anything, can we say about averages of the form (2.1.7) when T is a non-ergodic measure-preserving transformation of a probability space $(\Omega, \mathcal{F}, P)$? This question is of interest because often, when the underlying group Γ acts on a topological space $\mathcal{Y}$, there are random variables valued in $\mathcal{Y}$ whose orbits under *left* random walks on Γ form *stationary* but *non-ergodic* sequences on $\mathcal{Y}$.

Example 2.5.1 Let S^1 be the unit circle, viewed as a subset of the vector space $\mathbb{R}^2$, and let $\psi : S^1 \to S^1$ be the antipodal map $\psi(u) = -u$. Suppose that $\xi_1, \xi_2, \cdots$ is a sequence of random variables on the probability space $(\Omega, \mathcal{F}, P)$ taking values in the set $\{\psi, \mathrm{id}\}$, where $\mathrm{id} : S^1 \to S^1$ is the identity map, such that

$$P\{\xi_i = \psi\} = P\{\xi_i = \mathrm{id}\} = \frac{1}{2}. \tag{2.5.1}$$

Suppose also that U is an S^1-valued random variable on $(\Omega, \mathcal{F}, P)$ that is independent of $\xi_1, \xi_2, \cdots$ and has the uniform distribution (i.e., normalized arclength measure) on S^1. Then the sequence $\{\xi_n \xi_{n-1} \cdots \xi_2 \xi_1 \cdot U\}_{n \geq 0}$ is stationary, but not ergodic.

Exercise 2.5.2 Verify this, and show that the invariant σ-algebra $\mathcal{I}$ is generated by the random variable $V = \chi(U)$, where $\chi : S^1 \to \mathbb{RP}^1$ is the natural projection to real projective space.

Assume henceforth that T is a measure-preserving transformation of $(\Omega, \mathcal{F}, P)$, and let $\mathcal{I}$ be the σ-algebra of invariant events. A real-valued random variable Z is measurable with respect to $\mathcal{I}$ if and only if it is invariant, that is, $Z = Z \circ T$; clearly, for any invariant random variable Z,

$$\frac{1}{n} \sum_{i=0}^{n-1} Z \circ T^i = Z \quad \text{for every } n \in \mathbb{N}, \tag{2.5.2}$$

and so the relation (2.1.7) fails trivially for every nonconstant invariant random variable Z. Nevertheless, (2.5.2) shows that for every invariant random variable Z the ergodic averages converge: in particular,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} Z \circ T^i = Z. \tag{2.5.3}$$

At the other extreme from the invariant random variables are the *cocycles*. For any exponent $p \geq 1$, a cocyle of class L^p is a real random variable W of the form

$$W = Y - Y \circ T \quad \text{where} \quad Y \in L^p(\Omega, \mathcal{F}, P). \tag{2.5.4}$$

Any L^p-cocycle is also an L^1-cocycle, as Hölder's inequality implies that the L^p-norms $\|Y\|_p$ of any real random variable Y are nondecreasing in p. The L^2-cocycles are *orthogonal* to the invariant random variables of class L^2: in particular, if $W = Y - Y \circ T$ is an L^2-cocycle and $Z \in L^2$ is an invariant random variable, then

$$\begin{aligned} EWZ &= EYZ - E(Y \circ T)Z \\ &= EYZ - E(Y \circ T)(Z \circ T) \\ &= 0. \end{aligned}$$

Ergodic sums for a cocycle (2.5.4) telescope:

$$\sum_{i=0}^{n-1} W \circ T^i = Y - Y \circ T^n,$$

so the ergodic averages converge to 0:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} W \circ T^i = \lim_{n\to\infty} (Y - Y \circ T^n)/n = 0 \;\; \text{almost surely}, \tag{2.5.5}$$

the last by Corollary A.5.1 of the Appendix.

Proposition 2.5.3 *If T is a measure-preserving transformation of a probability space $(\Omega, \mathcal{F}, P)$, then every real random variable $X \in L^2(\Omega, \mathcal{F}, P)$ has an essentially unique decomposition*

$$X = Z + W \tag{2.5.6}$$

where Z is invariant and W is an L^2-cocycle. Consequently, for every random variable $X \in L^2(\Omega, \mathcal{F}, P)$ the ergodic averages converge:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i = Z \quad \textit{almost surely.} \tag{2.5.7}$$

The random variable Z is the L^2-orthogonal projection of W onto the closed linear subspace of invariant random variables, and W is the orthogonal projection onto the linear subspace of L^2-cocycles. See [111], Chapter 10, Section 8 for the basic theory of Hilbert spaces, and in particular Problem 53 for the essential facts about orthogonal projections. For any probability space $(\Omega, \mathcal{F}, P)$, the corresponding L^2-space is a real Hilbert space with inner product

$$\langle X, Y \rangle = E(XY).$$

Specializing to the probability space $(\Omega, \mathcal{I}, P)$ shows that the set of square-integrable invariant random variables is a *closed* linear subspace of $L^2(\Omega, \mathcal{F}, P)$; thus, the orthogonal projection mapping $X \mapsto Z$ is well-defined.

Proof. The convergence of the ergodic averages (2.5.7) will follow immediately from the decomposition (2.5.6), in view of (2.5.3) and (2.5.5). To establish the existence and uniqueness of the decomposition (2.5.6), we must show that the orthogonal complement in $L^2(\Omega, \mathcal{F}, P)$ of the space $L^2(\Omega, \mathcal{I}, P)$ of invariant random variables is the space of cocycles. We have already seen that every invariant L^2-random variable is orthogonal to the space of L^2-cocycles; thus, it remains to show that *every* L^2-random variable orthogonal to all cocycles is an invariant random variable.

By (2.1.5), the mapping $Y \mapsto Y \circ T$ is an isometry of L^2. Consequently, for any random variable $X \in L^2(\Omega, \mathcal{F}, P)$, the linear functional

$$Y \mapsto EX(Y \circ T)$$

is well-defined for all $Y \in L^2(\Omega, \mathcal{F}, P)$, and is norm-bounded: in particular, by the Cauchy-Schwarz inequality,

$$|EX(Y \circ T)| \le \|X\|_2 \, \|Y \circ T\|_2 \le \|X\|_2 \, \|Y\|_2 \, .$$

Therefore, by the Riesz Representation Theorem for Hilbert spaces (cf. [111], Section 10.8, Proposition 28) there exists an essentially unique random variable $\tilde{X} \in L^2(\Omega, \mathcal{F}, P)$ such that

$$EX(Y \circ T) = E\tilde{X}Y \quad \text{for all } Y \in L^2(\Omega, \mathcal{F}, P). \tag{2.5.8}$$

Suppose now that $X \in L^2(\Omega, \mathcal{F}, P)$ is orthogonal to every L^2-cocycle. Then for every $Y \in L^2(\Omega, \mathcal{F}, P)$ we have

$$EX(Y \circ T) = EXY \quad \Longrightarrow \quad EXY = E\tilde{X}Y,$$

whence it follows that $\tilde{X} = X$ almost surely. Using the fact that the identity (2.5.8) holds for $Y = X$, we conclude that

$$\begin{aligned}
\|X - X \circ T\|_2^2 &= \langle X - X \circ T, X - X \circ T \rangle \\
&= \langle X, X \rangle + \langle X \circ T, X \circ T \rangle - 2 \langle X, X \circ T \rangle \\
&= 2 \langle X, X \rangle - 2 \langle X, X \circ T \rangle \\
&= 2 \langle X, X \rangle - 2 \left\langle \tilde{X}, X \right\rangle
\end{aligned}$$

$$= 2\langle X, X\rangle - 2\langle X, X\rangle = 0,$$

so the random variable X must be (after change on an event of probability 0) invariant. This proves that the orthogonal complement of the space of cocycles is the space of invariant random variables. □

The convergence (2.5.7) of ergodic averages implies that the orthogonal projection $X \mapsto Z$ is a *positive* operator: if $X \geq 0$ then $Z \geq 0$. Since orthogonal projection is also linear, it follows that if $X_1 \leq X_2$ then the corresponding orthogonal projections Z_1, Z_2 satisfy $Z_1 \leq Z_2$. This, together with Wiener's Maximal Inequality (Proposition 2.4.2), allows us to extend the almost sure convergence relation (2.5.7) to all integrable random variable X, as follows. Assume first that $X \geq 0$ is nonnegative, and for each positive integer m define $X_m = \min(X, m)$ to be the level-m truncation of X. Since for any $m \in \mathbb{N}$ the random variable X_m is bounded, it is an element of L^2, and so by (2.5.7), for each $m \in \mathbb{N}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X_m \circ T^i = Z_m \quad \text{almost surely,} \tag{2.5.9}$$

where Z_m is the orthogonal projection of X_m onto the subspace $L^2(\Omega, \mathcal{I}, P)$ of invariant random variables. Since $0 \leq X_1 \leq X_2 \leq \cdots$, we also have that $0 \leq Z_1 \leq Z_2 \leq \cdots$, and so

$$Z :=\uparrow \lim_{m\to\infty} Z_m \tag{2.5.10}$$

exists. Moreover, since $X_m \leq X$, it follows that

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i \geq Z \quad \text{almost surely.} \tag{2.5.11}$$

On the other hand, the Monotone Convergence Theorem implies that $EX_m \uparrow EX$, so for any scalar $\varepsilon > 0$ there exists $m \in \mathbb{N}$ such that $EX - EX_m < \varepsilon^2$. Hence, by the Maximal Inequality,

$$P\left\{\sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{i=0}^{n-1} (X \circ T^i - X_m \circ T^i) > \varepsilon\right\} \leq \varepsilon.$$

This together with (2.5.9) implies that for each $m \in \mathbb{N}$,

$$P\left\{\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i - Z_m \geq \varepsilon\right\} \leq \varepsilon,$$

and since $\varepsilon > 0$ is arbitrary, it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i \leq Z \quad \text{almost surely.} \tag{2.5.12}$$

Relations (2.5.11) and (2.5.12) together prove that for any nonnegative, integrable random variable X the ergodic averages $n^{-1} \sum_{i=0}^{n-1} X \circ T^i$ converge almost surely, and the relation (2.5.10) provides an indirect description of the limit random variable Z. Following is a more useful characterization.

Proposition 2.5.4 *The random variable Z defined by equation* (2.5.10) *is the unique nonnegative, invariant random variable such that for every invariant event A,*

$$E(Z\mathbf{1}_A) = E(X\mathbf{1}_A). \tag{2.5.13}$$

Proof. Nonnegative random variables $Z \in L^1(\Omega, \mathcal{I}, P)$ are uniqeuly determined by their integrals on events $A \in \mathcal{I}$, by Example A.2.5, so there can be at most one random variable Z satisfying (2.5.13) up to changes on events of probability 0.

To prove that (2.5.13) holds for the random variable Z defined by (2.5.10), note first that if $A \in \mathcal{I}$ then $\mathbf{1}_A \in L^2(\Omega, \mathcal{I}, P)$. Since for each $m \in \mathbb{N}$ the random variable Z_m is the orthogonal projection of X_m onto the subspace $L^2(\Omega, \mathcal{I}, P)$, it follows that for any invariant event A and every $m \in \mathbb{N}$,

$$EX_m\mathbf{1}_A = EZ_m\mathbf{1}_A.$$

But $X = \uparrow \lim_{m\to\infty} X_m$ and $Z = \uparrow \lim_{m\to\infty} Z_m$, so two applications of the Monotone Converegence Theorem yield

$$\begin{aligned} EX\mathbf{1}_A &= \lim_{m\to\infty} EX_m\mathbf{1}_A \\ &= \lim_{m\to\infty} EZ_m\mathbf{1}_A \\ &= EZ\mathbf{1}_A \quad \text{for every } A \in \mathcal{I}. \end{aligned}$$

□

Proposition 2.5.4 shows that the limit random variable Z is the *conditional expectation* of X given the invariant σ-algebra $\mathcal{I}$.

Definition 2.5.5 Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}$ be a σ-algebra contained in $\mathcal{F}$. For any real-valued random variable $X \in L^1(\Omega, \mathcal{F}, P)$, the *conditional expectation* of X given $\mathcal{G}$ is the essentially unique integrable, $\mathcal{G}$-measurable random variable Z such that

$$E(X\mathbf{1}_G) = E(Z\mathbf{1}_G) \quad \text{for every } G \in \mathcal{G}. \tag{2.5.14}$$

This random variable is denoted by $E(X \mid \mathcal{G})$.

In the special case where $\mathcal{G} = \mathcal{I}$, we have proved the *existence* of conditional expectations for nonnegative random variables; for arbitrary real-valued random variables X , existence follows by linearity of the expectation operator, using the decomposition $X = X_+ - X_-$ into positive and negative parts, and setting

$$E(X \mid \mathcal{I}) = E(X_+ \mid \mathcal{I}) - E(X_- \mid \mathcal{I}). \tag{2.5.15}$$

Existence of conditional expectations for arbitrary σ-algebras $\mathcal{G}$ can be handled by the same argument. (Existence can also be proved using the Radon-Nikodym Theorem: see Section A.9 for this, along with a more complete discussion of the basic properties of conditional expectation.) Finally, essential uniqueness follows because integrable random variables are uniquely determined by their expectations on events of the underlying σ-algebra — see Example A.2.5 in Section A.2 of the Appendix.

Theorem 2.5.6 (Birkhoff's Ergodic Theorem) *Let T be a measure-preserving transformation of a probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{I}$ be the σ-algebra of invariant events. Then for any random variable $X \in L^1(\Omega, \mathcal{F}, P)$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ T^i = E(X \mid \mathcal{I}) \quad \textit{almost surely}, \tag{2.5.16}$$

where $E(X \mid \mathcal{I})$ is the conditional expectation of X given the σ-algebra $\mathcal{I}$.

Proof. For nonnegative random variables X this follows immediately from relations (2.5.11) and (2.5.12). The general case follows directly using the decomposition $X = X_+ - X_-$ into positive and negative parts, by (2.5.15), since

$$\sum_{i=0}^{n-1} X \circ T^i = \sum_{i=0}^{n-1} X_+ \circ T^i - \sum_{i=0}^{n-1} X_- \circ T^i.$$

□

Additional Notes. Another approach to the general form of Birkhoff's Theorem (Theorem 2.5.6) is via the *Ergodic Decomposition Theorem*. This states that if the measurable space $(\Omega, \mathcal{F})$ is a *Borel space* – for example, if Ω is a separable metric space and $\mathcal{F} = \mathcal{B}_\Omega$ is the σ-algebra of Borel sets – then every T-invariant probability measure on $\mathcal{F}$ is a *mixture* of *ergodic* invariant probability measures. Specifically, there is a Borel probability measure ν on the space $\mathcal{E}$ of ergodic invariant probability measures such that

$$P = \int_{\mathcal{E}} Q \, d\nu(Q). \tag{2.5.17}$$

This theorem is, in turn, an easy consequence of the existence of *regular conditional measures* on Borel spaces. See, e.g., Einsiedler and Ward [37], Theorem 6.2.

Birkhoff's Theorem has engendered a vast stream of extensions and generalizations. See Krengel [81] and Templeman [123] for extended suveys. Ergodic theorems for actions of *amenable groups* (see Chapter 5 for the definition) by measure-preserving transformations have been obtained under a variety of hypotheses; see Lindenstrauss [90] for some of the best results in this direction. Ergodic theorems for actions of *nonamenable* groups, such as lattices of semisimple Lie groups, have proven to be important in a variety of dynamical and number-theoretic contexts; see, for instance, Gorodnik and Nevo [52, 53] for lucid accounts of some of these. For more on cut points of random walks in the integer lattices $\mathbb{Z}^d$, see Lawler [86].

Chapter 3
Subadditivity and Its Ramifications

3.1 Speed of a Random Walk

Every random walk on a finitely generated group Γ whose step distribution has a finite first moment — that is, such that $E|X_1| < \infty$, where $|\cdot|$ is the word metric norm — travels at a definite *speed* (possibly 0). This, as we will show, is a consequence of the *subadditivity* and *invariance* of the word metric, and therefore generalizes to *any* (left-)invariant metric on the group, that is, a metric d such that

$$d(x, y) = d(gx, gy) \quad \text{for all } \; g, x, y \in \Gamma. \tag{3.1.1}$$

Theorem 3.1.1 *If d is an invariant metric on Γ, then for any random walk X_n on Γ such that $Ed(1, X_1) < \infty$,*

$$\lim_{n\to\infty} \frac{d(X_n, 1)}{n} = \inf_{n\geq 1} \frac{Ed(X_n, 1)}{n} \quad \textit{almost surely.} \tag{3.1.2}$$

In Section 3.5 we will prove a much more general theorem, the *Subadditive Ergodic Theorem*, from which Theorem 3.1.1 follows. In the special case where d is the word metric, we set

$$\ell := \inf_{n\geq 1} \frac{E|X_n|}{n}, \tag{3.1.3}$$

and call this constant the *speed* of the random walk. The speed depends, of course, on the choice of generating set $\mathbb{A}$.

For any symmetric random walk $(S_n)_{n\geq 0}$ on the integer lattice $\mathbb{Z}^d$ whose step distribution has finite first moment, the speed is 0. This follows easily from the usual Strong Law of Large Numbers, as symmetry implies that the mean vector of the step distribution is $\mathbf{0}$. Thus, the SLLN implies that for each of the d coordinates

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_3

$S_n^{(i)}$ of the random walk, $S_n^{(i)}/n \to 0$, and so the distance from the origin grows sublinearly.

It is not always true that a symmetric random walk on a finitely generated group has speed 0 — the isotropic nearest-neighbor random walk on $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$, as we have seen, has speed $\frac{1}{3}$. This prompts an obvious question: for which symmetric random walks, on which groups, is the speed 0, and for which is it positive? One of our main goals in these lectures will be to answer this question, at least partially.

3.2 Subadditive Sequences

A sequence a_n of real numbers is said to be *subadditive* if for every pair m, n of indices,

$$a_{m+n} \leq a_n + a_m.$$

Lemma 3.2.1 (Subadditivity Lemma) *For any subadditive sequence a_n, the limit $\lim_{n\to\infty} a_n/n$ exists and equals $\alpha = \inf_{n\geq 1} a_n/n \geq -\infty$.*

Proof. Exercise. □

Example 3.2.2 Recall that for a finitely generated group Γ, the ball $\mathbb{B}_n$ of radius n in with center at the group identity 1 consists of those group elements x such that $|x| \leq n$, where $|\cdot|$ is the word norm (relative to the generating set $\mathbb{A}$). Each ball $\mathbb{B}_n$ contains at most $|\mathbb{A}|^n$ elements,and by the triangle inequality, for any integers $m, n \geq 0$,

$$|\mathbb{B}_{n+m}| \geq |\mathbb{B}_n||\mathbb{B}_m|.$$

Thus, the sequence $-\log |\mathbb{B}_n|$ is subadditive, and so the limit

$$\beta := \lim_{n\to\infty} \frac{\log |\mathbb{B}_n|}{n} \tag{3.2.1}$$

exists and is nonnegative. The quantity β is the *exponential growth rate* of the group relative to the generating set $\mathbb{A}$. Clearly, $0 \leq \beta \leq \log |\mathbb{A}|$.

Exercise 3.2.3 The exponential growth rate β will generally depend on the choice of the generating set. Show, however, that if the exponential growth rate is 0 for one finite, symmetric generating set, then it is 0 for every such generating set.

Exercise 3.2.4 † It is easily checked that the exponential growth rate of $\mathbb{Z}^d$ is 0, while the exponential growth rate of the free group $\mathbb{F}_d$, relative to the standard generating set, is $\log(2d-1)$. What is the exponential growth rate of the lamplighter group $\mathbb{Z} \wr \mathbb{Z}_2^{\mathbb{Z}}$?

NOTE: There are two natural generating sets for the lamplighter group. The exponential growth rate will depend on which generating set is used.

Associated with any random walk X_n on any group Γ are a number of important subadditive sequences. In the following examples, the entries of the subadditive sequence a_n are obtained by taking expectations of functions w_n of the first n entries of the sequence $\xi_1, \xi_2, \cdots$ of steps of the random walk. This pattern presents itself sufficiently often that it deserves a name.

Definition 3.2.5 A sequence of functions $w_n : \mathcal{Y}^n \to \mathbb{R} \cup \{-\infty\}$ is *subadditive* if for any sequence $y_1, y_2, y_3, \cdots \in \mathcal{Y}$ and each pair m, n of nonnegative integers,

$$w_{n+m}(y_1, y_2, \cdots, y_{m+n}) \leq w_m(y_1, y_2, \cdots y_m) + w_n(y_{m+1}, y_{m+2}, \cdots, y_{m+n}). \tag{3.2.2}$$

The sequence $w_n : \mathcal{Y}^n \to [0, \infty)$ is *submultiplicative* if the sequence $\log w_n$ is subadditive (with the convention that $\log 0 = -\infty$).

If w_n is a subadditive sequence of functions and if $\mathbf{Y} = (Y_n)_{n\geq 1}$ is a stationary sequence of $\mathcal{Y}$-valued random variables such that $Ew_1(Y_1)_+ < \infty$ then the sequence

$$a_n = Ew_n(Y_1, Y_2, \cdots, Y_n)$$

is a subadditive sequence of real numbers, and so $\lim a_n/n = \inf a_n/n$. If $(f_n)_{n\geq 1}$ is a submultiplicative sequence of functions and if the random variables Y_n are independent and identically distributed , then the sequence of expectations $b_n = Ef_n(Y_1, \cdots, Y_n)$ is submultiplicative, because

$$\begin{aligned} Ef_{n+m}(Y_1, Y_2, \cdots, Y_{n+m}) &\leq Ef_n(Y_1, Y_2, \cdots, Y_n) f_m(Y_{n+1}, Y_{n+2}, \cdots, Y_{n+m}) \\ &= Ef_n(Y_1, Y_2, \cdots, Y_n) Ef_m(Y_1, Y_2, \cdots, Y_m). \end{aligned}$$

This need not hold if the sequence $\mathbf{Y} = (Y_n)_{n\geq 1}$ is merely stationary.

Example 3.2.6 For any invariant metric d on a group Γ, the sequence $w_n : \Gamma^n \to [0, \infty)$ defined by $w_n(y_1, y_2, \cdots, y_n) = d(y_1 y_2 \cdots y_n, 1)$ is subadditive. Consequently, for any random walk $X_n = \xi_1 \xi_2 \cdots \xi_n$ on Γ, if $Ed(X_1, 1) < \infty$ then

$$\ell(d) := \lim_{n\to\infty} \frac{Ed(X_n, 1)}{n} \tag{3.2.3}$$

exists and satisfies $0 \leq \ell(d) < \infty$. If d is the word metric on a finitely generated group Γ and the random walk is nearest-neighbor, then $\ell = \ell(d) \leq 1$.

Remark 3.2.7 Because $d(X_n, 1) = d(1, X_n^{-1})$, the relation (3.2.3) implies that the random walk $(X_n)_{n\geq 0}$ has the same speed as the *reflected random walk*

$$\hat{X}_n := \xi_1^{-1}\xi_2^{-1}\cdots\xi_n^{-1}, \tag{3.2.4}$$

that is, the random walk $(\hat{X}_n)_{n\geq 0}$ whose steps are the inverses of the steps ξ_i of $(X_n)_{n\geq 0}$.

Exercise 3.2.8 Show that a random walk has positive speed relative to an invariant metric d if and only if every lazy variant has positive speed relative to d.

HINT: Recall that a lazy variant of a random walk with step distribution μ is a random walk with step distribution $r\mu + (1-r)\delta_1$ for some $0 < r < 1$, where δ_1 is the unit point mass at the group identity 1. Show that for any $0 < r < 1$ and $n \in \mathbb{Z}_+$,

$$(r\mu + (1-r)\delta_1)^{*n} = \sum_{m=0}^{n} \binom{n}{m} r^m \mu^{*m}.$$

Example 3.2.9 The word metric (i.e., the usual graph distance in the Cayley graph) is the first that comes to mind, but it is not the only invariant metric of interest in random walk theory. For any *symmetric, transient* random walk on a finitely generated group Γ there is a second metric, specific to the step distribution, called the *Green metric*, that measures distance by the log of the hitting probability. For each $x \in \Gamma$, define

$$u(x) = P\{X_n = x \text{ for some } n \geq 0\};$$

then $u(xy) \geq u(x)u(y)$ for any elements $x, y \in \Gamma$, because the random walker can reach xy by first going to x and then from x to xy, so the function

$$d_G(x, y) := -\log u(x^{-1}y)$$

is a metric. (The hypothesis of *symmetry* guarantees that $d_G(x, y) = d_G(y, x)$. Symmetry also implies that d_G is a *metric*, and not merely a pseudo-metric, that is, $u(x) < 1$ for any $x \neq 1$. This is because if $u(x) = u(x^{-1}) = 1$ then the probability of return to the group identity 1 would be one, which would contradict the hypothesis of transience.) The limit

$$\gamma := \lim_{n\to\infty} \frac{Ed_G(X_n, 1)}{n} \tag{3.2.5}$$

is the *Green speed* of the random walk.

Example 3.2.10 Let $X_n = \xi_1\xi_2\cdots\xi_n$ be a random walk on a finitely generated group Γ with period d. (Recall that the period of a random walk is the greatest common divisor of the set $\{n \geq 1 : p_n(1, 1) > 0\}$ of return probabilities.) Define functions $f_n : \Gamma^n \to \{0, 1\}$ by

$$\begin{aligned} f_n(y_1, y_2, \cdots, y_n) &= 1 \quad \text{if } y_1 y_2 \cdots y_n = 1, \\ &= 0 \quad \text{if } y_1 y_2 \cdots y_n \neq 1. \end{aligned}$$

The sequence f_{nd} is submultiplicative, and hence so is the sequence of expectations $E f_{nd}(\xi_1, \xi_2, \cdots, \xi_n)$, which are the return probabilities $p_{dn}(1, 1)$. Consequently, the limit

$$\varrho := \lim_{n\to\infty} p_{dn}(1, 1)^{1/dn} \leq 1 \tag{3.2.6}$$

exists. This limit ϱ is called the *spectral radius* of the random walk; it will be studied in Chapter 5.

Example 3.2.11 Denote by μ the step distribution of a random walk $(X_n)_{n\geq 0}$ on a finitely generated group Γ and by μ^{*n} its nth convolution power, that is, $\mu^{*n}(x) = P\{X_n = x\}$. Define functions $f_n : \Gamma^n \to [0, 1]$ by

$$f_n(x_1, x_2, \cdots, x_n) = \mu^{*n}(x_1 x_2 \cdots x_n).$$

This sequence is submultiplicative, and so the sequence $w_n = \log f_n$ is subadditive. Therefore, if

$$h := h(\Gamma; \mu) = \inf_{n\geq 1} \frac{-E \log \mu^{*n}(X_n)}{n} < \infty \tag{3.2.7}$$

then

$$h = \lim_{n\to\infty} \frac{-E \log \mu^{*n}(X_n)}{n}. \tag{3.2.8}$$

The quantity $h = h(\Gamma; \mu)$ is called the *Avez entropy* of the random walk; it will be studied in depth in Chapter 10.

Remark 3.2.12 The *Shannon entropy* of a probability distribution μ on a countable set Θ is defined by

$$H(\mu) := -\sum_{x\in\Theta} \mu(x) \log \mu(x). \tag{3.2.9}$$

The Avez entropy h of a random walk with step distribution μ is the limit $\lim_{n\to\infty} H(\mu^{*n})/n$; this exists and is finite if and only if for some $n \geq 1$ the nth convolution power μ^{*n} has finite Shannon entropy.

Exercise 3.2.13 Show that for any *symmetric* random walk,

$$h \geq -\log \varrho. \tag{3.2.10}$$

HINT: Symmetry implies that $P\{X_{2n} = 1\} \geq P\{X_n = x\}^2$, for any $x \in \Gamma$.

Exercise 3.2.14 Show that if the speed ℓ and Avez entropy h are both finite, then they satisfy the basic inequality

$$h \leq \beta \ell, \tag{3.2.11}$$

where β is the exponential growth rate of the group. This holds for any invariant metric on Γ, not just the word metric. One of the interesting open problems in the subject is to characterize those groups for which equality can hold in this relation.

Exercise 3.2.15 Show that the Green speed γ (cf. Example 3.2.9) satisfies the inequality

$$\gamma \leq h. \tag{3.2.12}$$

HINT: The probability that the random walk $(X_n)_{n \geq 0}$ *ever* vists the site x is no smaller than the probability that it visits x at a specified time n.

Note 3.2.16 In Chapter 5 we will see that the oppositive inequality also holds, that is, $\gamma = h$. This result is due to Blachere, Haissinsky, and Mathieu [12]; see also Benjamini and Peres [9].

3.3 Kingman's Subadditive Ergodic Theorem

Theorem 3.3.1 (Subadditive Ergodic Theorem) *If $Y_1, Y_2, \cdots$ is a stationary, ergodic sequence of random variables taking values in a (measurable) space $\mathcal{Y}$, and if $w_n : \mathcal{Y}^n \to \mathbb{R}$ is a subadditive sequence of measurable functions such that*

$$E(w_1(Y_1))_+ < \infty \tag{3.3.1}$$

then with probability one,

$$\lim_{n\to\infty} \frac{w_n(Y_1, Y_2, \cdots, Y_n)}{n} = \inf_{m \geq 1} \frac{E w_m(Y_1, Y_2, \cdots, Y_m)}{m} := \alpha. \tag{3.3.2}$$

Here x_+ and x_- are the positive and negative parts of x, that is, $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$. Any sequence of independent, identically distributed random variables is stationary and ergodic (Example 2.1.2), so (3.3.2) holds, in particular, for the sequence $Y_n = \xi_n$, where ξ_n are the increments of a random walk.

The proof of Theorem 3.3.1 will be given in Section 3.5. Here are some easy consequences.

Corollary 3.3.2 *Let $(X_n)_{n\geq 0}$ be a random walk on a finitely generated group Γ whose step distribution μ has finite first moment*

$$E|X_1| = \sum_{x\in\Gamma} |x|\mu(x), \tag{3.3.3}$$

where $|\cdot|$ is the word length norm relative to some finite, symmetric generating set $\mathbb{A}$. Then with probability one,

$$\lim_{n\to\infty} \frac{|X_n|}{n} := \ell := \inf_{n\geq 1} \frac{E|X_n|}{n}. \tag{3.3.4}$$

Corollary 3.3.3 *Let $(X_n)_{n\geq 0}$ be a random walk on a finitely generated group Γ whose step distribution μ has finite Avez entropy h. Then with probability one,*

$$\lim_{n\to\infty} -\log\mu^{*n}(X_n)/n = h := \inf_{m\geq 1}(-E\log\mu^{*n}(X_n)/n). \tag{3.3.5}$$

As we have already observed, the word metric is not always the only metric of interest. When the group Γ is a *matrix group*, that is, when (say) $\Gamma \subset GL(k,\mathbb{R})$,the group of all invertible $k\times k$ matrices with real entries, the metric determined by the log of the *matrix norm*

$$\|M\| := \max_{v\in\mathbb{R}^k\,:\,\|v\|=1} \|Mv\| \tag{3.3.6}$$

is equally important. (Note: For a vector $v\in\mathbb{R}^d$ in a finite-dimensional Euclidean space, the notation $\|v\|$indicates the Euclidean length of v.) Like Corollary 3.3.2, the following important theorem is a direct consequence of the subadditive law of large numbers.

Corollary 3.3.4 (Furstenberg–Kesten [44]) *If $Y_1, Y_2, \cdots$ are independent, identically distributed random matrices taking values in $GL(k,\mathbb{R})$ such that*

$$E\log\|Y_1\|_+ < \infty \tag{3.3.7}$$

then with probability one,

$$\lim_{n\to\infty} \|Y_1Y_2\cdots Y_n\|^{1/n} := e^{\alpha} \quad \textit{where} \quad \alpha := \inf_{n\geq 1} E\log\|Y_1Y_2\cdots Y_n\|/n. \tag{3.3.8}$$

3.4 The Trail of a Lamplighter

Lamplighter random walks over the base group $\mathbb{Z}$ — that is, random walks on the wreath product group $\mathbb{Z}_2 \wr \mathbb{Z}$ — were introduced in Section 1.7, where the rationale for the name was explained. In this section, we consider lamplighter random walks over arbitrary finitely generated base groups Γ; these are by definition random walks on restricted wreath products $\mathbb{L} = \mathbb{Z}_2 \wr \Gamma$. In general, wreath products are defined as follows.

Definition 3.4.1 Let Γ be a finitely generated group. The *restricted configuration group* over Γ is the additive group $\oplus_\Gamma \mathbb{Z}_2$ of finitely supported functions $\psi : \Gamma \to \mathbb{Z}_2$ (that is, functions such that $\{x \in \Gamma : \psi(x) \neq 0\}$ is finite); addition is done coordinatewise modulo 2, that is,

$$(\psi_1 + \psi_2)(g) := \psi_1(g) + \psi_2(g) \mod 2 \quad \text{for } g \in \Gamma. \tag{3.4.1}$$

The *restricted wreath product* $\mathbb{L} = \mathbb{L}_\Gamma = \mathbb{Z}_2 \wr \Gamma$ is the finitely generated group whose elements are ordered pairs (x, ψ), where $x \in \Gamma$ and $\psi \in \oplus_\Gamma \mathbb{Z}_2$, with multiplication defined by the rule

$$(x, \psi) * (y, \varphi) = (xy, \psi + \sigma^x \varphi). \tag{3.4.2}$$

Here xy is the group product in Γ, and $\sigma^x \psi$ is the translated configuration whose value at $y \in \Gamma$ is $\sigma^x \psi(y) = \psi(x^{-1}y)$. The group identity in $\mathbb{L}$ is the pair $(1, \mathbf{0})$, where 1 is the identity in Γ and $\mathbf{0}$ is the configuration with all sites x set to $0 \in \mathbb{Z}_2$.

The projection mapping $\mathrm{pr} : \mathbb{L} \to \Gamma$ onto the first coordinate of a wreath product is a group homomorphism; consequently, if $(X_n)_{n \geq 0} = (S_n, L_n)_{n \geq 0}$ is a random walk on $\mathbb{L}$ with initial point $(1, \mathbf{0})$, then $(S_n)_{n \geq 0}$ is a random walk on Γ with initial point 1. Henceforth, we will refer to the group Γ as the *base group*, and the projection $(S_n)_{n \geq 0}$ as the *base random walk*.

Proposition 3.4.2 *Let $(X_n)_{n \geq 0} = (S_n, L_n)_{n \geq 0}$ be a lamplighter random walk on $\mathbb{L} = \mathbb{Z}_2 \wr \Gamma$ whose step distribution has finite support. If the base random walk $(S_n)_{n \geq 0}$ is transient, then with probability one the lamp configuration stabilizes as $n \to \infty$, i.e., for every $x \in \Gamma$,*

$$\lim_{n \to \infty} L_n(x) := L_\infty(x) \tag{3.4.3}$$

exists almost surely.

Exercise 3.4.3 Prove this.

Hint: Since the base random walk $(S_n)_{n \geq 0}$ is transient, Exercise 1.4.13 implies that with probability one, $\lim_{n \to \infty} |S_n| = \infty$. Consequently, since the step distribution of the random walk $(X_n)_{n \geq 0}$ has finite support, eventually the random walker will no longer be in range to flip the state of the lamp at any fixed $x \in \Gamma$.

Theorem 3.4.4 (Gilch [51]) *Let $(X_n)_{n\geq 0} = (S_n, L_n)_{n\geq 0}$ be a lamplighter random walk on $\mathbb{L} = \mathbb{Z}_2 \wr \Gamma$ whose step distribution has finite support. If the base random walk $(S_n)_{n\geq 0}$ is transient, then there is a constant $\beta \geq 0$ such that with probability one*

$$\lim_{n\to\infty} \frac{|support(L_n)|}{n} = \beta. \tag{3.4.4}$$

exists almost surely. Furthermore, if the random walk is irreducible then $\beta > 0$.

Here $\text{support}(L_n) := \{x \in \Gamma : L_n(x) = 1\}$ is the set of sites whose lamps are turned *on* after n. Thus, the content of the theorem is that for an irreducible lamplighter random walk the number of lamps turned on at time n grows linearly with n. Clearly, irreducibility is needed for this: if, for instance, the random walker never flips the state of any lamp, then $\text{support}(L_n) = \emptyset$ with probability 1, for every $n \in \mathbb{N}$, and so $\beta = 0$.

Corollary 3.4.5 *Under the hypotheses of Theorem 3.4.4, if the random walk $(X_n)_{n\geq 0}$ is irreducible then it has positive speed.*

Proof. Irreducibility implies that the constant β is positive. Consequently, by (3.4.4), the number of lamps turned on after n steps grows linearly with n. But since the step distribution of the random walk has finite support, the number of lamps that can be turned on or off in one step is bounded; therefore, for large n the distance from the initial state $(1, \mathbf{0})$ must also grow linearly with n. $\square$

We will deduce the almost sure convergence (3.4.4) from the Subadditive Ergodic Theorem. The positivity of the limit β requires a separate argument; the key step is the following lemma.

Lemma 3.4.6 *Under the hypotheses of Theorem 3.4.4, there exists $m \in \mathbb{N}$ such that for both $i = 0$ and $i = 1$,*

$$p_i := P\{L_n(1) = i \text{ for all } n \geq m\} > 0. \tag{3.4.5}$$

Proof. By hypothesis, the step distribution of the lamplighter random walk is finite, so there is a finite integer K such that at any time n the random walker cannot disturb the states of lamps at sites farther than K steps away from the walker's current location S_n. Furthermore, the base random walk is transient, so $\lim_{m\to\infty} P\{|S_n| > K \ \ \forall\, n \geq m\} = 1$, by Exercise 1.4.13. Consequently, there must be an element $x \in \Gamma$ such that if the random walker reaches site x then his chance of returning to the ball of radius K is less than $1/2$, that is,

$$P^x\{|S_n| > K \ \ \forall\, n \geq 0\} \geq \frac{1}{2}. \tag{3.4.6}$$

Now the full lamplighter random walk is irreducible, so for some $m, m' \in \mathbb{N}$ the probabilities $p := P\{X_m = (x, \mathbf{0})\}$ and $p' := P\{X_{m'} = (x, \delta_1)\} > 0$ are positive.

(Here $\mathbf{0}$ denotes the lamp configuration with all lamps off, and δ_1 the configuration with the lamp at the group identity $1 \in \Gamma$ on and all others off.) Thus, by (3.4.6),

$$P\left\{L_n(0) = 0 \text{ for all } n \geq \max(m, m')\right\} \geq \frac{p}{2} > 0 \quad \text{and}$$

$$P\left\{L_n(0) = 1 \text{ for all } n \geq \max(m, m')\right\} \geq \frac{p'}{2} > 0.$$

□

Proof of Theorem 3.4.4 We will prove the theorem only for lamplighter random walks $(S_n, L_n)_{n\geq 0}$ with initial point $(1, \mathbf{0})$; the general case then follows by routine arguments (which the reader is encouraged to supply). Let $\xi_1, \xi_2, \cdots$ be the increments of the random walk, and for any integer $m \geq 0$ define $L_{n;m}$ to be the lamp configuration after n steps of the random walk with increments $\xi_{m+1}, \xi_{m+2}, \cdots$ and initial state $(X_m, \mathbf{0})$; thus, $L_{n;m}$ would be the lamp configuration at time $n+m$ if all lamps were reset to 0 after the first m steps of the random walk. The group product law (3.4.2) implies that

$$L_{n+m} = L_m + L_{n;m} \mod 2 \quad \Longrightarrow$$
$$|\text{support}(L_{n+m})| \leq |\text{support}(L_m)| + |\text{support}(L_{n;m})|.$$

This shows that the sequence $Y_n := |\text{support}(L_n)|$ is subadditive. Since the step distribution has finite support, there exists $K \in \mathbb{N}$ such that the lamplighter can alter the states of at most K lamps at any step; consequently, for any $n \in \mathbb{N}$ the random variable Y_n is bounded by nK, and so

$$EY_1 = E|\text{support}(L_1)| < \infty.$$

Therefore, the Subadditive Ergodic Theorem implies that the limit (3.4.4) exists with probability one and is constant. Since the random variables Y_n/n are bounded, it follows that the convergence also holds in L^1, that is,

$$\lim_{n\to\infty} \frac{EY_n}{n} = \beta. \tag{3.4.7}$$

It remains to show that if the random walk is irreducible then the limit constant β is positive. For this it suffices to prove that $\liminf_{n\to\infty} \frac{EY_n}{n} > 0$. If a site x is visited at some time $k \leq n - m$, where m is as in Lemma 3.4.6, and if $L_{n'+k}(x) = 1$ (mod 2) for all $n' \geq k + m$, then x will be one of the sites counted in Y_n. On the event $\{S_k = x\} \cap \{L_k(x) = i\}$, the event that $L_{n'+k}(x) = 1 \mod 2$ for all $n' \geq k + m$ coincides with the event that $L_{n';k} = 1 + i \mod 2$ for all $n' \geq m$; hence,

$$Y_n \geq \sum_{k=0}^{n-m} \mathbf{1}_{G_{k,i}} \mathbf{1}_{F_{k,i}}$$

where

$$G_{k,i} = \{S_k \notin \{1, S_1, S_2, \cdots, S_{k-1}\}\} \cap \{L_k(S_k) = i\} \quad \text{and}$$
$$F_{k,i} = \left\{L_{n';k}(S_k) = 1 - i \mod 2 \text{ for all } n' \geq m\right\}.$$

For any integer $k \geq 0$ and $i = 0$ or $i = 1$, the event $F_{k,i}$ depends only on the steps of the random walk after time k, while $G_{k,i}$ depends only on the steps of the random walk up to time k; therefore, the events $G_{k,i}$ and $F_{k,i}$ are independent. Moreover, for every $k \in \mathbb{Z}_+$,

$$P(F_{k,0}) = p_1 \quad \text{and} \quad P(F_{k,1}) = p_0,$$

where p_0, p_1 are the constants defined by (3.4.5), which by Lemma 3.4.6 are both positive. Consequently,

$$2EY_n \geq \min(p_0, p_1) \sum_{k=0}^{n-m} E\mathbf{1}_{G_k}$$

where $G_k = G_{k,0} \cup G_{k,1}$ is the event $\{S_k \notin \{1, S_1, S_2, \cdots, S_{k-1}\}\}$ that the site S_k has not been visited previously by the base random walk. But

$$\sum_{k=0}^{n-m} \mathbf{1}_{G_k} = R_{n-m}$$

is the number of distinct sites visited by time $n-m$, and by Theorem 2.2.2, the ratio R_n/n converges almost surely to the probability that the base random walk never returns to its initial site. Since the base random walk is transient, this probability is positive. Therefore, by another application of the dominated convergence theorem, we have

$$\liminf_{n\to\infty} \frac{2EY_n}{n} \geq \min(p_0, p_1) P\{S_n \neq 1 \text{ for any } n \in \mathbb{N}\} > 0.$$

$\square$

3.5 Proof of Kingman's Theorem†

Proof of the Upper Bound. First, we will prove the easier half of the theorem, that with probability 1,

$$\limsup_{m\to\infty} \frac{w_m(Y_1, Y_2, \cdots, Y_m)}{m} \leq \alpha. \tag{3.5.1}$$

For notational ease, let's write

$$W(m; n) = w_n(Y_{m+1}, Y_{m+2}, \cdots, Y_{m+n}). \tag{3.5.2}$$

Iteration of the subadditivity relation (3.2.2) shows that for any partition of an interval $[0, k]$ of the integers into nonoverlapping subintervals $[m_i, m_i + n_i)$ (ordered so that $m_{i+1} = m_i + n_i$),

$$W(0; k) \leq \sum_i W(m_i; n_i); \tag{3.5.3}$$

consequently, for any integers $m, n \geq 1$ and $k \geq 0$,

$$W(0; mn + k) \leq \sum_{i=0}^{m-1} W(in; n) + W(mn; k)_+. \tag{3.5.4}$$

Applying this with $n = 1$ and $k = 0$ yields $W(0; m)_+ \leq \sum_{1\leq i\leq m} w_1(Y_i)_+$, and so it follows from the hypothesis (3.3.1) that $EW(0; m)_+ < \infty$. Now for any fixed $n \geq 0$, the random variables $(W(m; n)_+)_{m\geq 0}$ are identically distributed, since the underlying sequence $(Y_n)_{n\geq 0}$ is stationary; consequently, by Corollary A.5.1 of the Appendix, for any $k \geq 1$,

$$\lim_{m\to\infty} W(m; k)_+/m = 0 \quad \text{almost surely.} \tag{3.5.5}$$

The crucial fact is that for each choice of $n \geq 1$ the sequence of random variables $(W(in; n))_{i=0,1,\cdots}$ that occur in the sum (3.5.4) is stationary and ergodic, in particular, these random variables have the same form (2.1.8) as the summands in Birkhoff's Ergodic Theorem (Theorem 2.1.6), with $f = w_n$. Thus, by (2.1.8) and (3.5.5), for each $n \geq 1$, with probability one,

$$\begin{aligned}\limsup_{m\to\infty} \max_{0\leq k<n} &\frac{W(0; mn + k)}{mn} \\ &\leq \limsup_{m\to\infty} (mn)^{-1} \left\{ \sum_{i=0}^{m-1} W(in; n) + \max_{k\leq n} W(mn; k)_+ \right\}\end{aligned}$$

$$= \lim_{m\to\infty} (mn)^{-1} \sum_{i=0}^{m-1} W(in; n)$$

$$= \frac{EW(0; n)}{n}.$$

Since n is arbitrary, the upper bound (3.5.1) follows. □

The proof of the lower bound is more elaborate; it will require the following preliminary result.

Lemma 3.5.1 *Under the hypotheses of Theorem 3.3.1, the random variable*

$$\liminf_{n\to\infty} \frac{w_n(Y_1, Y_2, \cdots, Y_n)}{n}$$

is almost surely constant.

This in turn will rely on the following elementary fact.

Lemma 3.5.2 *Let X and Y be real-valued random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$. Suppose that $X \leq Y$ almost surely, and that X and Y have the same distribution, i.e., for every $z \in \mathbb{R}$,*

$$P\{X \leq z\} = P\{Y \leq z\}.$$

Then $X = Y$ almost surely.

Proof. Exercise. □

Proof of Lemma 3.5.1 Let $W(m; n)$ be defined by equation (3.5.2); then by subadditivity and (3.5.5), for every integer $m \geq 1$,

$$\liminf_{n\to\infty} \frac{W(0; n)}{n} \leq \liminf_{n\to\infty} \frac{W(0; m) + W(m; n)}{n} = \liminf_{n\to\infty} \frac{W(m; n)}{n}.$$

For each $m \geq 1$, the sequence $(W(m; n))_{n\geq 0}$ has the same joint distribution as $(W(0; n))_{n\geq 0}$, so the liminfs on the left and right have the same distribution; thus, by Proposition 3.5.2, they are almost surely equal. Therefore,

$$\liminf_{n\to\infty} \frac{W(0; n)}{n} = \limsup_{m\to\infty} \liminf_{n\to\infty} \frac{W(m; n)}{n}$$

almost surely. The random variable on the right is invariant (that is, it is measurable with respect to the invariant σ-algebra $\mathbf{Y}^{-1}(\mathcal{I})$: see Section 2.1). But by hypothesis, the sequence $\mathbf{Y} = (Y_n)_{n\geq 1}$ is *ergodic*, so the invariant σ-algebra is trivial (every invariant event has probability 0 or 1); consequently, every invariant random variable is almost surely constant. □

Proof of the Lower Bound This will recapitulate the proof of Wiener's Maximal Inequality in Section 2.4; once again, the Covering Lemma 2.4.1 will play a central role. As in the proof of the upper bound, we use the abbreviation (3.5.2), and we set $w = w_1$. By Lemma 3.5.1, there exists a constant $\beta \in \mathbb{R} \cup \{-\infty\}$ such that with probability 1,

$$\liminf_{n\to\infty} \frac{W(0;n)}{n} = \beta. \tag{3.5.6}$$

By inequality (3.5.1), we have $\beta \leq \alpha$; to complete the proof we must show that $\beta \geq \alpha$.

Fix a real number $\gamma > \beta$ and an integer $K \geq 1$, and for each integer $m \geq 0$ define the event

$$G_m = G_{m,K} = \left\{ \min_{1\leq n\leq K} \frac{W(m;n)}{n} > \gamma \right\}.$$

Since the driving sequence $(Y_n)_{n\geq 1}$ is stationary, the indicators $\mathbf{1}_{G_m}$ are identically distributed, as are the random variables $w(Y_{m+1})_+\mathbf{1}_{G_m}$. Furthermore, since (3.5.6) holds with probability one, $\lim_{K\to\infty} P(G_{0;K}) = \lim_{K\to\infty} P(G_{m,K}) = 0$. Hence, by the first moment hypothesis (3.3.1) and the dominated convergence theorem,

$$\lim_{K\to\infty} Ew(Y_1)_+\mathbf{1}_{G_{0;K}} = \lim_{K\to\infty} Ew(Y_{m+1})_+\mathbf{1}_{G_{m,K}} = 0. \tag{3.5.7}$$

For each point $\omega \in \Omega$ of the underlying probability space, define $L = L(\omega)$ to be the set of all nonnegative integers m such that $\omega \in G_m^c$; then by definition of the event G_m, for each $m \in L$, there is an integer n such that $1 \leq n \leq K$ and

$$W(m;m+n) \leq n\gamma. \tag{3.5.8}$$

Define $\mathcal{J}(\omega)$ to be the set of all intervals $[m, m+n)$ with $n \leq K$ for which (3.5.8) holds. By definition, $L(\omega)$ is the set of all left endpoints of intervals in $\mathcal{J}(\omega)$. Consequently, by the Covering Lemma 2.4.1, there is a subset $\mathcal{J}_*(\omega)$ of $\mathcal{J}(\omega)$ consisting of nonoverlapping intervals $J = [m, m+n)$ such that every element of $L = L(\omega)$ is contained in one of the intervals $J \in \mathcal{J}_*(\omega)$.

Assume that the intervals $J_i = [m_i, m_i + n_i)$ in the collection $\mathcal{J}_*(\omega)$ are indexed from left to right, that is, so that $m_i + n_i \leq m_{i+1}$. By construction, each interval J_i has length $n_i \leq K$, and for each integer m not in one of the intervals J_i it must be the case that $\omega \in G_m$. Hence, by the subadditivity relation (3.5.3), for each $0 \leq j < K$ and any $n \geq 1$,

$$W(0;nK+j) \leq \sum_{i\,:\,m_i\leq nK-K} W(m_i;n_i) + \sum_{m=0}^{nK-1} w(Y_{m+1})_+\mathbf{1}_{G_m} + \sum_{i=nK-K+1}^{nK+j} w(Y_i)_+.$$

For each interval $J_i = [m_i, m_i + n_i)$ the contribution of $W(m_i; n_i)$ to the sum is no greater than $\gamma n_i = \gamma |J_i|$, by (3.5.8); therefore, because the intervals J_i are pairwise disjoint, we have

$$W(0; nK + j) \leq \gamma nK + \sum_{m=0}^{nK-1} w(Y_{m+1})_+ \mathbf{1}_{G_m} + \sum_{i=nK-K+1}^{nK+j} w(Y_i)_+.$$

Taking expectations, dividing by nK, and letting $n \to \infty$ yields

$$\alpha := \lim_{n\to\infty} \frac{EW(0; n)}{n} \leq \gamma + Ew(Y_1)_+ \mathbf{1}_{G_{0,K}}.$$

This holds for every integer $K \geq 1$. Since $\lim_{K\to\infty} Ew(Y_1)_+ \mathbf{1}_{G_{0,K}} = 0$, it follows that $\alpha \leq \gamma$ for every $\gamma > \beta$. This proves that $\alpha \leq \beta$. □

Additional Notes. The *Avez entropy* of a random walk was introduced by Avez in his article [4]. It will be the subject of Chapter 10 below. The result of Exercise 3.2.14 is known as the *fundamental inequality*; it is due to Guivarc'h [61]. Recently, Gouëzel, Mathéus, and Maucourant [57] have proved that the fundamental inequality is strict for any symmetric random walk whose step distribution has finite support in any nonelementary hyperbolic group that is not virtually free. (See Chapter 13 for the relevant definitions.)

Theorem 3.3.1 is due to Kingman [78]. It is an essential tool in modern probability theory. In addition to Kingman's article [78], see the survey article [79], with discussion by several authors, some of which give important examples of its use, and others outline different approaches to the proof. See also Steele [119] for an elegant alternative proof.

The Furstenberg-Kesten Theorem (Corollary 3.3.4) is the first step toward a deep and important theorem describing the evolution of random matrix products due to Oseledec [105]. See Walters [130] for a nice proof, and Filip [40] for an interesting modern approach based on earlier work of Karlsson and Ledrappier [72] Ruelle [113] and Karlsson and Margulis [75] give some infinite-dimensional generalizations of the Furstenberg-Kesten and Oseledec Theorems.

Chapter 4
The Carne-Varopoulos Inequality

4.1 Statement and Consequences

N. Varopoulos [128] discovered a remarkable upper bound for the n-step transition probabilities of a symmetric, nearest-neighbor random walk on a finitely generated group. Shortly afterward, K. Carne [22] found an elegant approach to Varopoulos' inequality that yielded (as was noted later by R. Lyons) a simpler and even sharper bound. Lyons' improvement of the Carne-Varopoulos inequality can be stated as follows.

Theorem 4.1.1 (Carne-Varopoulos) *Let $(X_n)_{n\geq 0}$ be a symmetric, nearest-neighbor random walk on a finitely generated group Γ with word metric d and norm $|x| := d(1, x)$. Let $\varrho = \lim P^1\{X_{2n} = 1\}^{1/2n}$ be the spectral radius of the random walk. Then for every element $x \in \Gamma$ and every $n \geq 1$,*

$$P^1\{X_n = x\} \leq 2\varrho^n \exp\{-|x|^2/2n\} \tag{4.1.1}$$

The hypothesis of symmetry is crucial. Consider, for instance the random walk on $\mathbb{Z}$ that at each step moves $+1$ to the right: for this random walk,

$$P^0\{X_n = n\} = 1.$$

The most basic case of the Carne-Varopoulos inequality is the simple random walk on the integer lattice $\mathbb{Z}$. Here the spectral radius is $\varrho = 1$ (by Stirling's formula), so (4.1.1) follows from the *Hoeffding inequality* (Proposition A.6.1 of the Appendix):

$$P\{X_n = x\} \leq P\{X_n \geq |x|\} \leq 2\exp\{-|x|^2/2n\}.$$

Ultimately, this is the origin of the generalization (4.1.1), as we will see.

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_4

Before turning to the proof of Theorem 4.1.1 (see Theorem 4.4.9 and Proposition 4.4.10 in Section 4.4), let's look at some of its implications.

Corollary 4.1.2 *If $(X_n)_{n\geq 0}$ is a symmetric, nearest-neighbor random walk on a finitely generated group Γ with speed ℓ and Avez entropy h, then*

$$h \geq \ell^2/2 - \log \varrho \tag{4.1.2}$$

Proof. By Corollary 3.3.2, for any $\varepsilon > 0$ and all sufficiently large n the probability will be nearly 1 that the norm $|X_n|$ will lie in the range $n\ell \pm n\varepsilon$. By the Carne-Varopoulos inequality, for any group element x at distance $n\ell \pm n\varepsilon$ from the group identity 1,

$$P\{X_n = x\} \leq 2\varrho^n \exp\left\{-|x|^2/2n\right\} \leq 2\varrho^n \exp\left\{-n(\ell-\varepsilon)^2/2\right\}.$$

Thus, with probability approaching 1 as $n \to \infty$,

$$\mu^{*n}(X_n) \leq 2\varrho^n \exp\left\{-n(\ell-\varepsilon)^2/2\right\}.$$

Since $\varepsilon > 0$ is arbitrary, the inequality (4.1.2) follows from Corollary 3.3.3 □

Corollary 4.1.2 implies that any symmetric, nearest-neighbor random walk with spectral radius $\varrho < 1$ must have positive Avez entropy. It also implies that a symmetric, nearest-neighbor random walk can have positive speed only if it also has positive Avez entropy. The next result shows that the converse is also true.

Proposition 4.1.3 *For any random walk $(X_n)_{n\geq 0}$ on a finitely generated group, if $(X_n)_{n\geq 0}$ has finite speed ℓ and finite Avez entropy h then*

$$h > 0 \implies \ell > 0.$$

Proof. A finitely generated group Γ has at most exponential growth (cf. Example 3.2.2): in particular, there is a constant $\beta \geq 0$ such that for large n the cardinality of the ball $\mathbb{B}_n$ of radius n satisfies $\log|\mathbb{B}_n| \sim n\beta$. If a random walk X_n on Γ has speed $\ell = 0$, then for any $\varepsilon > 0$ and all sufficiently large n the distribution of the random variable X_n will be almost entirely concentrated in $\mathbb{B}_{\varepsilon n}$, by Corollary 3.3.2. But since $\mathbb{B}_{\varepsilon n}$ contains at most $e^{2\beta\varepsilon n}$ distinct group elements (for large n), it follows that with probability approaching 1,

$$\mu^{*n}(X_n) \geq \exp\{-2\beta\varepsilon n\}.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $h = 0$. □

An easy modification of this argument (exercise!) shows that on groups of sub-exponential growth *all* symmetric random walks with finitely supported step distributions have entropy 0 and speed 0.

Proposition 4.1.4 *If a finitely generated group* Γ *has exponential growth rate* $\beta = 0$ *then every nearest-neighbor random walk on* Γ *has entropy* $h = 0$*, and consequently, every symmetric, nearest-neighbor random walk on* Γ *has speed* $\ell = 0$*.*

□

In Chapter 5, we will identify a large class of groups of exponential growth — the *nonamenable* groups — in which all symmetric, irreducible random walks have positive Avez entropies. In view of Proposition 4.1.4, this might lead one to suspect that positivity of the Avez entropy for symmetric random walks is completely determined by the geometry of the ambient group. This is false: the lamplighter group $\mathbb{Z}_2 \wr \mathbb{Z}$ supports symmetric random walks with speed and entropy 0 (cf. Section 1.7) and also symmetric random walks with positive entropy. See the fascinating article [43] for the rather intricate construction.

4.2 Chebyshev Polynomials

Definition 4.2.1 The nth Chebyshev polynomial of the first kind is the unique nth degree polynomial $T_n(x)$ with real (actually, integer) coefficients such that

$$T_n(\cos\theta) = \cos(n\theta). \tag{4.2.1}$$

The first few are

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x. \end{aligned}$$

For notational convenience, let's extend the definition to negative indices $m \in \mathbb{Z}$ by setting $T_{-m}(x) = T_m(x)$; since cos is an even function, the identity (4.2.1) is then valid for all $n \in \mathbb{Z}$. Note for later reference that the defining identity (4.2.1) implies that

$$|T_n(x)| \leq 1 \quad \text{for every} \quad x \in [-1, 1]. \tag{4.2.2}$$

Exercise 4.2.2 Take 30 seconds to convince yourself that polynomials satisfying the identities (4.2.1) exist and are unique. If you want to know more, check the WIKIPEDIA article on Chebyshev polynomials.

The appearance of Chebyshev polynomials in a series of lectures devoted to random walks on groups has a bit of the rabbit-out-of-a-hat character. However,

as the next result shows, there is a natural connection between the Chebyshev polynomials and simple random walk on $\mathbb{Z}$.

Proposition 4.2.3 *Let $(S_n)_{n\geq 0}$ be the simple random walk on $\mathbb{Z}$; then for every $n = 0, 1, 2, \cdots$*

$$x^n = \sum_{m\in\mathbb{Z}} P\{S_n = m\}\, T_m(x). \tag{4.2.3}$$

Proof. This is simply a reformulation of the fact that the characteristic function (i.e., Fourier transform) of the random variable S_n is the nth power of the characteristic function of S_1. In detail: first, writing $S_n = \sum_{j=1}^n \xi_j$, we have

$$\begin{aligned} Ee^{i\theta S_n} &= E\exp\left\{i\theta\sum_{j=1}^n \xi_j\right\} \\ &= \prod_{j=1}^n E\exp\left\{i\theta\xi_j\right\} \\ &= (Ee^{i\theta\xi_1})^n \\ &= (\cos\theta)^n \end{aligned}$$

This uses the fact that the increments ξ_j are independent and identically distributed. Next, express the same characteristic function as a sum over possible values of S_n:

$$\begin{aligned} Ee^{i\theta S_n} &= \sum_{m\in\mathbb{Z}} e^{i\theta m} P\{S_n = m\} \\ &= \sum_{m\in\mathbb{Z}} \frac{1}{2}(e^{+i\theta m} + e^{-i\theta m}) P\{S_n = m\} \\ &= \sum_{m\in\mathbb{Z}} \cos(m\theta) P\{S_n = m\} \\ &= \sum_{m\in\mathbb{Z}} T_m(\cos\theta) P\{S_n = m\}. \end{aligned}$$

Here, in the second equality, we have used the symmetry of the simple random walk, which implies that $P\{S_n = +m\} = P\{S_n = -m\}$. It now follows that for every real θ,

$$(\cos\theta)^n = \sum_{m\in\mathbb{Z}} T_m(\cos\theta) P\{S_n = m\}.$$

This implies that the polynomials on the left and right sides of equation (4.2.3) agree at infinitely many values of x, and so they must in fact be the same polynomial.

□

4.3 A Transfer Matrix Inequality

Since (4.2.3) is a polynomial identity, one can substitute for the variable x any element of an algebra over the reals. In particular, (4.2.3) applies with $x = L$, where L is any linear operator on a real vector space V (i.e., a linear transformation $L : V \to V$). For such an operator,

$$L^n = \sum_{m \in \mathbb{Z}} T_m(L) P\{S_n = m\}. \tag{4.3.1}$$

This becomes useful when the linear operator L is *self-adjoint*, because in this case the *Spectral Theorem* for self-adjoint operators provides a handle on the operator (matrix) norms of the operators $T_m(L)$. Recall (cf. equation (3.3.6)) that the *norm* of a linear operator $L : V \to V$ on a finite-dimensional, real inner product space V with inner product $\langle \cdot, \cdot \rangle$ and associated vector norm $\|\cdot\|$ is defined by

$$\|L\| := \max_{v \in V : \|v\|=1} \|Lv\| = \max_{v \in V : v \neq 0} \frac{\langle Lv, Lv \rangle^{1/2}}{\langle v, v \rangle^{1/2}}. \tag{4.3.2}$$

A linear operator L on the inner product space V is *self-adjoint* if

$$\langle v, Lw \rangle = \langle Lv, w \rangle \quad \text{for all } \ v, w \in V. \tag{4.3.3}$$

Spectral Theorem for Self-Adjoint Operators. *For any self-adjoint linear operator $L : V \to V$ on a finite-dimensional real inner product space V of dimension D there exist a complete orthonormal basis $\{u_i\}_{i \leq D}$ of V (the* eigenvectors) *and a corresponding set $\{\lambda_i\}_{i \leq D}$ of real numbers (the* eigenvalues) *such that*

$$Lu_i = \lambda_i u_i \quad \text{for each } i \in [D]. \tag{4.3.4}$$

Consequently, for every positive integer n and every vector $v \in V$,

$$L^n v = \sum_{i=1}^{D} \langle v, u_i \rangle \lambda_i^n u_i \quad \text{and therefore} \tag{4.3.5}$$

$$\left\|L^n\right\| := \max_{\|v\| \leq 1} \left\|L^n v\right\| = \max_{i \leq D} |\lambda_i|^n. \tag{4.3.6}$$

See Herstein [65], Section 6.10 or Horn & Johnson [67], Section 2.5 for the proof, along with the rudimentary theory of inner product spaces. The spectral representation (4.3.5) extends to polynomial functions of L: in particular, for any polynomial $f(x)$ with real coefficients,

$$f(L)v = \sum_{i=1}^{D} \langle v, u_i \rangle\, f(\lambda_i) u_i \quad \text{for all } v \in V. \tag{4.3.7}$$

Exercise 4.3.1 Use (4.3.7) to show that if L is a self-adjoint operator of norm $\|L\| \leq 1$ then for any Chebyshev polynomial $T_m(x)$,

$$\|T_m(L)\| \leq 1. \tag{4.3.8}$$

Proposition 4.3.2 *Let V be a finite-dimensional real inner product space V of dimension D with orthonormal basis $\{\mathbf{e}_i\}_{i \in [D]}$, and let L be a self-adjoint linear operator on V. For each pair of indices $i, j \in [D]$ define*

$$d(i, j) = \min \left\{ n \in \mathbb{Z}_+ \,:\, \left\langle \mathbf{e}_i, L^n \mathbf{e}_j \right\rangle \neq 0 \right\} \tag{4.3.9}$$

or $d(i, j) = \infty$ if there is no such n. Then for every pair of indices $i, j \in [D]$ and every nonnegative integer n,

$$|\left\langle \mathbf{e}_i, L^n \mathbf{e}_j \right\rangle| \leq 2 \,\|L\|^n \exp\left\{ -d(i, j)^2 / 2n \right\}. \tag{4.3.10}$$

Before we move on to the proof, let's look more closely at the meaning of the "distance" $d(i, j)$ in (4.3.10). In a variety of interesting contexts — including random walk theory, as we will show in Section 4.4 — the matrix entries $L_{i,j} := \left\langle \mathbf{e}_i, L\mathbf{e}_j \right\rangle$ represent physical or mathematical quantities, such as transition probabilities, attached to the edges of a graph $G = ([D], \mathcal{E})$ with vertex set $[D]$ and edge set $\mathcal{E}$. Thus, L is a symmetric, $D \times D$ matrix whose entries $L_{i,j}$ are nonzero if and only if $\{i, j\} \in \mathcal{E}$. Such matrices L are sometimes called *transfer operators* or *transfer matrices* (cf., e.g., Stanley [118], Section 4.7 or Thompson [124], Section 5–4).

For any operator L satisfying the hypotheses of Proposition 4.3.2 there is a graph on the vertex set $[D]$ relative to which L is a transfer matrix, to wit, the graph G with edges $\{i, j\}$ such that $L_{i,j} := \left\langle \mathbf{e}_i, L\mathbf{e}_j \right\rangle \neq 0$. The matrix entries $L^n_{i,j} := \left\langle \mathbf{e}_i, L^n \mathbf{e}_j \right\rangle$ of a power of L can be interpreted as *sums over paths*, as follows. For each $n \in \mathbb{N}$ and any pair $i, j \in [D]$, let $\mathcal{P}_n(i, j)$ be the set of all paths $x_0, x_1, \cdots, x_n \in [D]^{n+1}$ of length n in G from $x_0 = i$ to $x_n = j$ (thus, each pair x_i, x_{i+1} of successive vertices must be an edge of the graph), and for each such path γ let

$$w(\gamma) = \prod_{i=1}^{n} L_{x_i, x_j}. \tag{4.3.11}$$

Then

$$L^n_{i,j} = \sum_{\gamma \in \mathcal{P}_n(i,j)} w(\gamma). \tag{4.3.12}$$

Exercise 4.3.3 Verify this.

The representation (4.3.12) shows that the matrix entry $L^n_{i,j}$ will be nonzero only if there is a path of length n in the graph from i to j. Thus, the graph distance between vertices $i, j \in [D]$ is dominated by the distance $d(i, j)$ in inequality (4.3.10), and so (4.3.10) remains valid when $d(i, j)$ is replaced by the graph distance. In the special case where the nonzero entries $L_{i,j}$ are *positive*, the graph distance coincides with the distance $d(i, j)$ in (4.3.10).

Proof of Proposition 4.3.2 Without loss of generality, we may assume that the operator L has norm $\|L\| = 1$, because multiplying L by a scalar does not affect the validity of the inequality (4.3.10). By (4.3.1), for any pair of indices i, j and any nonnegative integer n,

$$\left\langle \mathbf{e}_i, L^n \mathbf{e}_j \right\rangle = \sum_{m \in \mathbb{Z}} P\{S_n = m\} \left\langle \mathbf{e}_i, T_m(L) \mathbf{e}_j \right\rangle$$

where $\{S_n\}_{n \geq 0}$ is simple random walk on the integers. By definition $T_m(x)$ is a polynomial of degree $|m|$; since $\left\langle \mathbf{e}_i, L^k \mathbf{e}_j \right\rangle = 0$ for every $k < d(i, j)$, it follows that if $|m| < d(i, j)$ then $\left\langle \mathbf{e}_i, T_m(L) \mathbf{e}_j \right\rangle = 0$. Hence, we have

$$\left\langle \mathbf{e}_i, L^n \mathbf{e}_j \right\rangle = \sum_{|m| \geq d(i,j)} P\{S_n = m\} \left\langle \mathbf{e}_i, T_m(L) \mathbf{e}_j \right\rangle.$$

But by Exercise 4.3.1, for every integer m the operator $T_m(L)$ has matrix norm ≤ 1, and so $|\left\langle \mathbf{e}_i, T_m(L) \mathbf{e}_j \right\rangle| \leq 1$; therefore,

$$\begin{aligned} |\left\langle \mathbf{e}_i, L^n \mathbf{e}_j \right\rangle| &\leq \sum_{|m| \geq d(i,j)} P\{S_n = m\} \\ &= P\{|S_n| \geq d(i, j)\} \\ &\leq 2 \exp\left\{-d(i, j)^2/2n\right\}, \end{aligned}$$

the last by Hoeffding's inequality (A.6.1). ◻

4.4 Markov Operators

The (infinite) matrix $(p(x, y))_{x,y\in\Gamma}$ of transition probabilities of a nearest-neighbor random walk on a finitely generated group Γ has nonzero entries $p(x, y) > 0$ only for those pairs $(x, y) \in \Gamma \times \Gamma$ such that x, y are the endpoints of an edge of the Cayley graph of Γ. Thus, if the random walk is symmetric then the transition probability matrix satisfies all of the hypotheses of Proposition 4.3.2 except one: it is not finite-dimensional. In this section we will show that the Carne-Varopoulos inequality (4.1.1) can nevertheless be deduced from Proposition 4.3.2 by taking finite sections of the matrix $(p(x, y))_{x,y\in\Gamma}$.

Let Γ be a finitely generated group with finite, symmetric generating set $\mathbb{A}$. Denote by $C_b(\Gamma)$ the Banach space of bounded, real-valued functions on Γ with supremum norm $\|\cdot\|_\infty$, and by $L^2(\Gamma)$ the real Hilbert space of square-summable functions $f : \Gamma \to \mathbb{R}$ with norm and inner product

$$\|f\|_2^2 := \sum_{x\in\Gamma} f(x)^2 \quad \text{and} \quad \langle f, g\rangle := \sum_{x\in\Gamma} f(x)g(x). \tag{4.4.1}$$

(See Royden [111], Chapter 10, or Rudin [112], Chapters 4–5 for the fundamentals of Banach and Hilbert space theory.) If $p(\cdot, \cdot)$ are the transition probabilities of a random walk $(X_n)_{n\geq 0}$ on Γ, then for any function $f \in C_b(\Gamma)$ the function

$$\mathbb{M}f(x) := E^x f(X_1) = \sum_{y\in\Gamma} p(x, y)f(y) \tag{4.4.2}$$

is well-defined and bounded, with sup norm no larger than that of f; consequently, $\mathbb{M} : C_b(\Gamma) \to C_b(\Gamma)$ is a *(bounded) linear operator* on $C_b(\Gamma)$. Since $L^2(\Gamma) \subset C_b(\Gamma)$, the restriction of $\mathbb{M}$ to $L^2(\Gamma)$ is well-defined; in fact, this restriction maps n $L^2(\Gamma)$ into itself, by the following lemma.

Lemma 4.4.1 *For any function $f \in L^2(\Gamma) \subset C_b(\Gamma)$,*

$$\|\mathbb{M}f\|_2 \leq \|f\|_2 . \tag{4.4.3}$$

Exercise 4.4.2 Prove this.

HINT: First, show that it suffices to consider only *nonnegative* functions f. Next, show that $\langle \mathbb{M}f, \mathbb{M}f\rangle = \sum_{x\in\Gamma} \sum_{y\in\Gamma} \sum_{z\in\Gamma} f(xy)f(xz)p(1, y)p(1, z)$. Finally, use Cauchy-Schwarz to prove that for any $y, z \in \Gamma$,

$$\left|\sum_{x\in\Gamma} f(xy)f(xz)\right| \leq \|f\|_2^2 .$$

Lemma 4.4.1 implies that the restriction of $\mathbb{M}$ to $L^2(\Gamma)$ is a linear operator of norm $\|M\| \leq 1$; here the *(operator) norm* of a linear operator $T : L^2(\Gamma) \to L^2(\Gamma)$ is defined by

$$\|T\| := \sup_{f \in L^2(\Gamma) : \|f\|_2 = 1} \|Tf\|_2 . \tag{4.4.4}$$

If $\|T\| < \infty$ then T is said to a *bounded* linear operator; the mapping $T : L^2(\Gamma) \to L^2(\Gamma)$ induced by a bounded linear operator is continuous with respect to the L^2-metric.

Exercise 4.4.3 Show that if T is a bounded linear operator then the sequence $\log \|T^n\|$ is subadditive, that is, for any positive integers m, n,

$$\left\| T^{n+m} \right\| \leq \left\| T^n \right\| \left\| T^m \right\| . \tag{4.4.5}$$

This implies, by Lemma 3.2.1, that $\lim_{n\to\infty} \|T^n\|^{1/n}$ exists and is finite. The limit is called the *spectral radius* of the operator T.

Definition 4.4.4 The *Markov operator* of a random walk with transition probabilities $p(x, y)$ is the linear operator $\mathbb{M} : C_b(\Gamma) \to C_b(\Gamma)$, or its restriction $\mathbb{M} : L^2(\Gamma) \to L^2(\Gamma)$, defined by equation (4.4.2).

The Chapman-Kolmogorov equations (1.3.9) imply that if $\mathbb{M}$ is the Markov operator of a random walk on Γ then the matrix entries of $\mathbb{M}^n$ are the n-step transition probabilities, that is,

$$\mathbb{M}^n f(x) = E^x f(X_n) = \sum_{y \in \Gamma} p_n(x, y) f(y). \tag{4.4.6}$$

Consequently, the n-step transition probabilities can be recovered by

$$p_n(x, y) = \left\langle \delta_x, \mathbb{M}^n \delta_y \right\rangle . \tag{4.4.7}$$

Here $\delta_z(x) = \delta(z, x)$ is the Kronecker delta function on Γ, which takes the value 1 at $x = z$ and 0 elsewhere. A trivial calculation shows that if the random walk is *symmetric*, that is, if $p(x, y) = p(y, x)$ for all $x, y \in \Gamma$, then the Markov operator $\mathbb{M}$ is *self-adjoint*, that is, for all $f, g \in L^2(\Gamma)$,

$$\langle f, \mathbb{M}g \rangle = \langle \mathbb{M}f, g \rangle . \tag{4.4.8}$$

Exercise 4.4.5 Use equations (4.4.6) and (4.4.7) to show that the transition probabilities of a symmetric random walk satisfy

$$p_{2n}(x, y) \le p_{2n}(1, 1) \quad \text{for all} \quad n \in \mathbb{N} \quad \text{and} \quad x, y \in \Gamma.$$

HINT: $\left\langle \mathbb{M}^n(\delta_x - \delta_y), \mathbb{M}^n(\delta_x - \delta_y) \right\rangle \ge 0$.

Next, let's identify a useful sequence of finite-dimensional sections of the Markov operator $\mathbb{M}$. For each $k = 1, 2, \cdots$ let V_k be the finite-dimensional linear subspace of $L^2(\Gamma)$ consisting of those functions f which vanish outside the ball $\mathbb{B}_k$ of radius k in Γ centered at the group identify 1. Define Π_k to be the *orthogonal projection operator* for this subspace, that is, for each $g \in L^2(\Gamma)$, set

$$\begin{aligned} \Pi_k g(x) &= g(x) \quad \text{if} \quad x \in \mathbb{B}_k; \\ &= 0 \qquad\quad \text{if} \quad x \notin \mathbb{B}_k. \end{aligned}$$

Next, define $\mathbb{M}_k$ to be the restriction of $\mathbb{M}$ to V_k, that is,

$$\mathbb{M}_k = \Pi_k \mathbb{M}\ \Pi_k,$$

equivalently,

$$\mathbb{M}_k g(x) = \mathbf{1}_{\mathbb{B}_k}(x) \sum_{y \in \mathbb{B}_k} p(x, y) g(y). \tag{4.4.9}$$

The latter equation exhibits $\mathbb{M}_k : V_k \to V_k$ as the linear transformation with matrix representation $(p(x, y))_{x,y \in \mathbb{B}_k}$ relative to the natural basis $(\delta_x)_{x \in \mathbb{B}_k}$ for V_k. If the underlying random walk is symmetric then for each k the operator $\mathbb{M}_k$ is self-adjoint, as then the matrix $(p(x, y))_{x,y \in \mathbb{B}_k}$ will be real and symmetric. Therefore, for a symmetric random walk the operators $\mathbb{M}_k$ are subject to the conclusions of the Spectral Theorem for finite-dimensional self-adjoint operators.

Lemma 4.4.6 *If the random walk is symmetric then the norms of the operators* $\mathbb{M}_k$ *are nondecreasing in k, and*

$$\|\mathbb{M}\| = \lim_{k \to \infty} \|\mathbb{M}_k\|. \tag{4.4.10}$$

Exercise 4.4.7 Prove Lemma 4.4.6.

HINT: For the first assertion, the fact that the matrix entries (4.4.9) are nonnegative might be useful.

Exercise 4.4.8 Prove that $\|\mathbb{M}\| = \sup_{f \in L^2(\Gamma)} \langle f, \mathbb{M} f \rangle$.

HINT: The analogous statement for the operator $\mathbb{M}_k$ follows from the Spectral Theorem.

Theorem 4.4.9 *Let $p_n(\cdot, \cdot)$ be the n-step transition probabilities of a symmetric random walk on a finitely generated group Γ, and let $\mathbb{M}$ and $\mathbb{M}_k$ be the Markov and truncated Markov operators for the random walk. For any $x \in \Gamma$ and any $n \in \mathbb{Z}_+$, if $k \geq n$ then*

$$\begin{aligned} p_n(1, x) &\leq 2 \|\mathbb{M}_k\|^n \exp\left\{-|x|^2/2n\right\} \\ &\leq 2 \|\mathbb{M}\|^n \exp\left\{-|x|^2/2n\right\}, \end{aligned} \tag{4.4.11}$$

where $|x|$ is the word-length norm of x.

Proof. By equation (1.3.8), the n-step transition probability $p_n(1, x)$ is the sum of the path-probabilities over all paths of length n in the Cayley group from 1 to x. By equation (4.3.12), the matrix entry $\mathbb{M}_k^n(1, x)$ is the sum of the path-probabilities over all paths of length n in the Cayley group from 1 to x that do not exit the ball $\mathbb{B}_k$. Since no path of length n that starts at 1 can exit the ball $\mathbb{B}_n$, if $k \geq n$ then the two sums agree, and so

$$p_n(1, x) = \mathbb{M}_k^n(1, x) = \left\langle \delta_1, \mathbb{M}_k^n \delta_x \right\rangle.$$

Therefore, the first inequality in (4.4.11) follows directly from Proposition 4.3.2. The second inequality follows from Lemma 4.4.6. □

The Carne-Varopoulos inequality (4.1.1) is an immediate consequence of Theorem 4.4.9 and the following fact.

Proposition 4.4.10 *The norm of the Markov operator $\mathbb{M}$ of a symmetric random walk on a finitely generated group coincides with the spectral radius of the random walk, that is,*

$$\|M\| = \lim_{n \to \infty} p_{2n}(1, 1)^{1/2n} := \varrho. \tag{4.4.12}$$

Proof. The sub-multiplicativity of the operator norm (equation (4.4.5)) implies that $\|\mathbb{M}\| \geq \|\mathbb{M}^n\|^{1/n}$ for all integers $n \geq 1$, so to prove that $\|\mathbb{M}\| \geq \varrho$ it suffices to show that

$$\lim_{n \to \infty} \left\|\mathbb{M}^n\right\|^{1/n} \geq \lim_{n \to \infty} p_{2n}(1, 1)^{1/2n}. \tag{4.4.13}$$

But for any $x \in \Gamma$, the Kronecker delta function δ_x has L^2-norm 1, so by the definition (4.4.4) of the operator norm,

$$\left\|\mathbb{M}^n\right\|^2 \geq \left\langle \mathbb{M}^n \delta_1, \mathbb{M}^n \delta_1 \right\rangle = p_{2n}(1, 1).$$

This clearly implies (4.4.13), and hence also that $\|\mathbb{M}\| \geq \varrho$.

To prove the reverse inequality, it suffices, by Lemma 4.4.6, to show that for any $k \in \mathbb{N}$,

$$\|\mathbb{M}_k\| \le \varrho. \tag{4.4.14}$$

The operator $\mathbb{M}_k$ is, by definition, the restriction of the Markov operator to the finite-dimensional subspace V_k of functions supported by the ball $\mathbb{B}_k$ of radius k; consequently, by the Spectral Theorem (in particular, relations (4.3.5) and (4.3.6)), for any $n \in \mathbb{N}$ we have

$$\|\mathbb{M}_k\| = \left\|\mathbb{M}_k^n\right\|^{1/n} = \left\langle \mathbb{M}_k^n u_1, \mathbb{M}_k^n u_1 \right\rangle^{1/2n},$$

where u_1 is the normalized eigenvector of $\mathbb{M}_k$ corresponding to the lead eigenvalue $\lambda_1 = \|\mathbb{M}_k\|$. Thus, to prove the inequality (4.4.14) it is enough to show that for any $k \in \mathbb{N}$,

$$\limsup_{n\to\infty} \left\langle \mathbb{M}_k^n u_1, \mathbb{M}_k^n u_1 \right\rangle^{1/2n} \le \lim_{n\to\infty} p_{2n}(1,1)^{1/2n}. \tag{4.4.15}$$

The Kronecker delta functions $(\delta_x)_{x\in\mathbb{B}_k}$ constitute an orthonormal basis of V_k, so u_1 is a linear combination of these:

$$u_1 = \sum_{x\in\mathbb{B}_k} a_x \delta_x.$$

Consequently, by the Cauchy-Schwarz inequality and the nonnegativity of the matrix coefficients of the Markov operator $\mathbb{M}$, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\left\langle \mathbb{M}_k^n u_1, \mathbb{M}_k^n u_1 \right\rangle &\le \sum_{x\in\mathbb{B}_k}\sum_{y\in\mathbb{B}_k} |a_x a_y| \left\|\mathbb{M}_k^n \delta_x\right\| \left\|\mathbb{M}_k^n \delta_y\right\| \\
&\le \sum_{x\in\mathbb{B}_k}\sum_{y\in\mathbb{B}_k} |a_x a_y| \left\|\mathbb{M}^n \delta_x\right\| \left\|\mathbb{M}^n \delta_y\right\| \\
&= \left(\sum_{x\in\mathbb{B}_k}\sum_{y\in\mathbb{B}_k} |a_x a_y|\right) \left\|\mathbb{M}^n \delta_1\right\|^2 \\
&= \left(\sum_{x\in\mathbb{B}_k}\sum_{y\in\mathbb{B}_k} |a_x a_y|\right) p_{2n}(1,1).
\end{aligned}$$

The inequality (4.4.15) follows directly. □

Exercise 4.4.11 Show that for asymmetric random walks the equality (4.4.12) need not hold, but that the *inequality*

$$\limsup_{n\to\infty} p_n(1,1)^{1/n} \le \lim_{n\to\infty} \left\|\mathbb{M}^n\right\|^{1/n} = \inf_{n\ge 1} \left\|\mathbb{M}^n\right\|^{1/n} \tag{4.4.16}$$

is always valid.

Additional Notes. Varopoulos [128] proved a similar but weaker inequality than (4.1.1) and used it to prove that a symmetric random walk admits nonconstant, bounded, μ-harmonic functions if and only if it has positive speed. This result will follow from Theorem 10.1.1 in Chapter 10 below. Carne [22] discovered the formula (4.2.3) connecting the transition probabilities of simple random walk on $\mathbb{Z}$ to Chebyshev polynomials, and deduced the improved form (4.1.1) of Varopoulos' inequality, minus the spectral radius factor ϱ^n; R. Lyons pointed out that this improvement follows from Carne's proof. Mathieu [96] has found an analogue of the Carne-Varopoulos inequality for non-reversible Markov chains, and Peyre [107] gives a "probabilistic" proof of the inequality. Inequality (4.1.2) is essentially due to Varopoulos [128].

Chapter 5
Isoperimetric Inequalities and Amenability

5.1 Amenable and Nonamenable Groups

As we have seen, for every symmetric random walk on a finitely generated group the probability of return to the initial point after $2n$ steps decays roughly like ϱ^{2n}, where $\varrho \leq 1$ is the *spectral radius* of the walk. When is it the case that this exponential rate ϱ is strictly less than 1? Kesten [77] discovered that the answer depends on a fundamental structural feature of the ambient group: *amenability*.

Let $G = (V, \mathcal{E})$ be a connected, locally finite graph. (*Locally finite* means that every vertex $v \in V$ is incident to at most finitely many edges; equivalently, no vertex has infinitely many nearest neighbors.) For any finite set $F \subset V$, define the *outer boundary* ∂F to be the set of all vertices $y \notin F$ such that y is a nearest neighbor of some vertex $x \in F$.

Definition 5.1.1 The *isoperimetric constant* $\iota(G)$ is defined by

$$\iota(G) := \inf_{F \text{ finite}} \frac{|\partial F|}{|F|}. \tag{5.1.1}$$

Definition 5.1.2 A finitely generated group Γ is *amenable* (respectively, *non-amenable*) if its Cayley graph $G_{\Gamma;\mathbb{A}}$ (relative to any finite, symmetric set of generators) has isoperimetric constant $\iota(G_{\Gamma;\mathbb{A}}) = 0$ (respectively, $\iota(G_{\Gamma;\mathbb{A}}) > 0$).

This definition is essentially due to Folner [41], who showed that non-positivity of the isoperimetric constant is equivalent to the existence of a bi-invariant Banach mean value on Γ.

Exercise 5.1.3 Show that if the isoperimetric constant is 0 for *some* finite set of generators then it is 0 for *every* finite set of generators.

HINT: It suffices to show that the isoperimetric constant is positive for a symmetric generating set $\mathbb{A}$ if and only if for any $n \geq 1$ it is also positive for the generating set

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_5

$\cup_{i=1}^{n} \mathbb{A}^i$, where $\mathbb{A}^i$ is the set of group elements that can be represented as products of length i in the generators $\mathbb{A}$.

Thus, amenability (or nonamenability) is a property of the group, not of the particular Cayley graph (i.e., generating set) chosen. For a nonamenable group, of course, the isoperimetric constants might be different for different Cayley graphs; but for every Cayley graph, the isoperimetric constant must be positive.

Exercise 5.1.4

(A) Show that the d-dimensional integer lattice $\mathbb{Z}^d$ is amenable.
(B) Show that the free group $\mathbb{F}_d$ on $d \geq 2$ generators is nonamenable.
(C) Show that the d-dimensional lamplighter group $\mathbb{L}^d$ is amenable.

HINT: For (C), consider the sets F_m consisting of all pairs (x, ψ) with $x \in \mathbb{Z}^d$ such that $|x| \leq m$ and $\psi \equiv 0$ outside $|x| \leq m$.

The distinction between amenable and nonamenable groups is of fundamental importance in representation theory and geometric group theory, in part because of the large number of equivalent criteria for amenability. (A useful necessary and sufficient condition is given by Proposition 11.2.7 in Chapter 11 below. See [106] for more.) The significance of amenability in random walk begins with the following celebrated theorem of Kesten [76, 77].

Theorem 5.1.5 (Kesten) *Every symmetric, nearest-neighbor random walk on a finitely generated, amenable group has spectral radius* 1*, and every* irreducible*, symmetric random walk on a finitely generated, nonamenable group has spectral radius* < 1.

It is not difficult to see (exercise!) that this implies that every symmetric, irreducible random walk on a finitely generated, amenable group has spectral radius 1.

The hypothesis of irreducibility is obviously necessary for the second assertion, because if the step distribution of the random walk has full support in an amenable subgroup then the random walk would have spectral radius 1, by the first assertion of the theorem. This would be the case, for example, if $\Gamma = \mathbb{F}_2$ is the free group on two generators a, b and μ is the step distribution $\mu(a) = \mu(a^{-1}) = \frac{1}{2}$.

The proof of Kesten's Theorem that we will present in Sections 5.3–5.6 relies heavily on symmetry. The theorem, as stated, does not extend to asymmetric random walks: for instance, the nearest-neighbor $p - q$ random walk on $\mathbb{Z}$ with $p > q$ has exponentially decaying return probabilities (by Stirling's formula). Nevertheless, Kesten's Theorem has the following partial extension.

Theorem 5.1.6 *Every irreducible random walk on a finitely generated, nonamenable group has spectral radius* < 1.

This is due to M. Day: see [30] for the proof.

By Exercise 3.2.13, the spectral radius and the Avez entropy of a symmetric random walk obey the inequality $h \geq -\log \varrho$, and by Exercise 3.2.14 the speed ℓ satisfies $h \leq \beta \ell$, where β is the exponential growth rate of the group. Thus, Kesten's theorem has the following consequences.

Corollary 5.1.7 *If the group Γ is finitely generated and nonamenable, then every irreducible, symmetric random walk on Γ whose step distribution μ has finite first moment $\sum_{g\in\Gamma}|g|\mu(g)$ and finite Shannon entropy $\sum_{g\in\Gamma}-\mu(g)\log\mu(x)$ must have positive Avez entropy, and hence also positive speed.*

There are, of course, amenable groups for which every irreducible, symmetric random walk with finitely supported step distribution has positive speed: the lamplighter groups $\mathbb{L}^d$ in dimensions $d \geq 3$, for instance, are amenable, but random walks on these groups always have positive speed, by Theorem 3.4.4.

Since any nearest-neighbor random walks on a group of subexponential growth must have Avez entropy 0, Corollary 5.1.7 implies the following fact.

Corollary 5.1.8 *Every finitely generated group Γ with subexponential growth (i.e., exponential growth rate $\beta = 0$) is amenable.*

If $\varphi : \Gamma \to \Gamma'$ is a group homomorphism, then any random walk $(X_n)_{n\geq 0}$ on Γ projects via φ to a random walk $(\varphi(X_n))_{n\geq 0}$ on Γ'. If the random walk $(X_n)_{n\geq 0}$ has spectral radius 1, then so must the projection, because for any $n \geq 0$,

$$P\{X_n = 1\} \leq P\{\varphi(X_n) = 1\}.$$

Therefore, Kesten's Theorem provides yet another test for nonamenability.

Corollary 5.1.9 *If $\varphi : \Gamma \to \Gamma'$ is a surjective homomorphism onto a nonamenable group Γ', then Γ must be nonamenable.*

There several other useful tests for amenability/nonamenability that do not require direct estimation of the isoperimetric constant. The next two exercises provide criteria that are, in some cases, relatively easy to check.

Exercise 5.1.10 Show that if Γ has a finitely generated, nonamenable subgroup H then it is nonamenable. Thus, in particular, if Γ has a subgroup isomorphic to the free group $\mathbb{F}_2$ on two generators then it is nonamenable, by Exercise 5.2.3.

HINTS: (A) Without loss of generality, assume that the generating set of H is contained in that of Γ. (B) Let $\{x_i H\}_{i\in S}$ be an enumeration of the left cosets of H; then every subset $F \subset \Gamma$ can be partitioned as $F = \cup_{i\in S} F \cap (x_i H)$.

The group $SL(2,\mathbb{Z})$ has a free subgroup, as we will show in Section 5.2, so it is nonamenable. Since $SL(2,\mathbb{Z})$ is a subgroup of $SL(n,\mathbb{Z})$, for any integer $n \geq 2$, it follows that $SL(n,\mathbb{Z})$ is also nonamenable.

Exercise 5.1.11 * Let Γ be a finitely generated group with an amenable subgroup H of finite index. Show that Γ is amenable.
NOTE: This can be proved using only the definition 5.1.2 and elementary arguments. We will give another proof, based on Kesten's Theorem, in Section 7.4.

5.2 Klein's Ping-Pong Lemma

One way to prove that a finitely generated group is nonamenable is to show that it has a free subgroup (cf. Exercise 5.1.10). A useful tool for this is the following fact, known as *Klein's Ping-Pong Lemma*.

Lemma 5.2.1 (Klein) *Let Γ be a group generated by $a^{\pm 1}, b^{\pm 1}$, where a, b are elements of infinite order. Suppose that Γ acts on a set X (cf. Definition 1.2.4), and that X contains nonempty, nonoverlapping subsets X_A, X_B such that*

$$\begin{aligned} a^n(X_B) &\subset X_A \quad \text{for every } \ n \in \mathbb{Z} \setminus \{0\}, \quad \text{and} \\ b^n(X_A) &\subset X_B \quad \text{for every } \ n \in \mathbb{Z} \setminus \{0\}. \end{aligned} \tag{5.2.1}$$

Then Γ is isomorphic to the free group $\mathbb{F}_2$ with generators $a^{\pm 1}, b^{\pm 1}$.

Exercise 5.2.2 Prove this.

HINT: There is a natural surjective homomorphism $\varphi : \mathbb{F}_2 \to \Gamma$ that maps each reduced word in the generators $a^{\pm 1}, b^{\pm 1}$ to the corresponding product in Γ. It suffices to show that the kernel of this homomorphism is the empty word $\emptyset$, equivalently, that no nontrivial reduced word

$$\begin{aligned} w_1 &= a^{m_1} b^{n_1} a^{m_2} b^{n_2} \cdots a^{m_k} b^{n_k} \quad \text{or} \\ w_2 &= b^{n_1} a^{m_1} b^{n_2} \cdots a^{m_{k-1}} b^{n_k} a^{m_k} \quad \text{or} \\ w_3 &= b^{n_1} a^{m_2} b^{n_2} \cdots a^{m_k} b^{n_k} \quad \text{or} \\ w_4 &= a^{m_1} b^{n_1} a^{m_2} b^{n_2} \cdots a^{m_k} \end{aligned}$$

is mapped to the identity $1 \in \Gamma$ by φ. Use the ping-pong hypothesis to show that $\varphi(w) \neq 1$ for any word of the form $w = w_3$ or $w = w_4$. For a word $w = w_1$ with $m_1 \geq 1$, the word $w' = awa^{-1}$ is of the form $w' = w_4$, so $\varphi(w') \neq 1$; this implies that $\varphi(w) \neq 1$. Similar arguments take care of the remaining cases.

Exercise 5.2.3 Let Γ_A and Γ_B be the subgroups

$$\begin{aligned} \Gamma_A &= \left\{ \begin{pmatrix} \pm 1 & \pm 2n \\ 0 & \pm 1 \end{pmatrix} \mid n \in \mathbb{Z} \right\} \quad \text{and} \\ \Gamma_B &= \left\{ \begin{pmatrix} \pm 1 & 0 \\ \pm 2n & \pm 1 \end{pmatrix} \mid n \in \mathbb{Z} \right\} \end{aligned}$$

of $PSL(2, \mathbb{Z})$. Check that both Γ_A and Γ_B are isomorphic to $\mathbb{Z}$. Then use the Ping-Pong Lemma to show that the subgroup generated by $\Gamma_A \cup \Gamma_B$ is isomorphic to $\mathbb{F}_2$.

HINT: $PSL(2,\mathbb{Z})$ acts on $\mathbb{R}^2/\{\pm\}$ by matrix multiplication. Let X_A and X_B be the nonoverlapping angular sectors $|x| > |y|$ and $|y| > |x|$, respectively.

Exercise 5.2.3 shows that the modular group $PSL(2,\mathbb{Z})$ is nonamenable. Since $PSL(2,\mathbb{Z})$ is a homomorphic image of $SL(2,\mathbb{Z})$, it follows by Corollary 5.1.9 that $SL(2,\mathbb{Z})$ is also nonamenable. As we have remarked earlier (without proof, but see [2]), the modular group is isomorphic to the free product $\mathbb{Z}_2 * \mathbb{Z}_3$; thus, the nonamenability of the modular group is actually a special case of the following fact.

Proposition 5.2.4 *If Γ_A and Γ_B are finite groups with elements of order at least 2 and 3, respectively, then the free product $\Gamma := \Gamma_A * \Gamma_B$ has a subgroup isomorphic to $\mathbb{F}_2$, and therefore is nonamenable.*

Exercise 5.2.5 † The group $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2$ is the *infinite dihedral group*. In this exercise you will show, among other things, that this group is amenable.

(A) Let Γ_A be the two-element group with generator a and Γ_B the two-element group generated by b. The free product $\Gamma = \Gamma_A * \Gamma_B \cong \mathbb{Z}_2 * \mathbb{Z}_2$ consists of all finite words in the letters a, b in which the letters alternate, that is, words in which every a is followed by a b and every b by an a. Let H be the cyclic subgroup generated by the word ab. Show that H is infinite, and hence isomorphic to $\mathbb{Z}$.
(B) Check that H is a normal subgroup of Γ, and show that the quotient group Γ/H is isomorphic to $\mathbb{Z}_2$.
(C) Conclude from Exercise 5.1.11 that Γ is amenable.
(D) Show that Γ is isomorphic to the automorphism group of the Cayley graph of $\mathbb{Z}$.

HINT: let $a, b : \mathbb{Z} \to \mathbb{Z}$ be the automorphisms

$$a(x) = 1 - x,$$
$$b(x) = -x.$$

Proof of Proposition 5.2.4 Let X be the set of all infinite reduced words of the form

$$w = a_1 b_1 a_2 b_2 \cdots \quad \text{or} \quad w = b_1 a_2 b_2 a_3 \cdots,$$

where $a_i \in \Gamma_A \setminus \{1\}$ and $b_i \in \Gamma_B \setminus \{1\}$. There is a natural action of Γ on X by left multiplication followed by reduction: for example, if $\Gamma_A = \{1, a\}$ and $\Gamma_B = \{1, b, b^2\}$ are the two- and three- element groups, then

$$(abab^2) \cdot (ababa \cdots) = abab^2ababa \cdots \quad \text{and}$$
$$(abab^2) \cdot (babab \cdots) = ab[ab^2ba]ab \cdots$$
$$= ab^2ab \cdots .$$

Fix elements $a \in \Gamma_A \setminus \{1\}$ and $b \in \Gamma_B \setminus \{1\}$ such that $b^2 \neq 1$, and set

$$g = ababa \quad \text{and} \quad h = bab.$$

A routine induction shows that for every integer $n \geq 1$,

$$g^n = aba \cdots aba, \quad \text{and} \quad h^n = ba \cdots ab, \tag{5.2.2}$$

so both g and h are elements of infinite order. Let X_A be the subset of X consisting of all infinite words with first letter $a^{\pm 1}$, and X_B the subset consisting of all infinite words with first letter $b^{\pm 1}$. These sets are nonempty and nonoverlapping, and the ping-pong hypothesis (5.2.1) holds by virtue of (5.2.2). □

5.3 Proof of Kesten's Theorem: Amenable Groups

Assumption 5.3.1 *Assume for the remainder of this chapter that $(X_n)_{n \geq 0}$ is a symmetric random walk on a finitely generated group Γ, and that the symmetric generating set $\mathbb{A}$ has been chosen so that the step distribution μ assigns positive probability to every element $a \in \mathbb{A}$. Denote by $\mathbb{M}$ the Markov operator and by $p_n(x, y)$ the n-step transition probabilities of the random walk.*

The spectral radius ϱ of a symmetric random walk coincides with the norm $\|\mathbb{M}\|$ of its Markov operator (cf. Proposition 4.4.10). Therefore, to prove Kesten's theorem it suffices to show that the Markov operator of a symmetric, irreducible, nearest neighbor random walk on a finitely generated group Γ has norm 1 if Γ is amenable, but has norm < 1 if Γ is nonamenable. Since the norm of the Markov operator never exceeds 1, to prove the first of these assertions it suffices to show that if Γ is amenable then there exist functions $f \in L^2(\Gamma)$ for which the ratios $\|\mathbb{M} f\|_2 \,/\, \|f\|_2$ come arbitrarily close to 1. For *nearest-neighbor* random walks, this is a consequence of the following lemma.

Lemma 5.3.2 *Assume that the step distribution of the random walk is supported by the generating set $\mathbb{A}$. Then for any nonempty, finite set $F \subset \Gamma$,*

$$\frac{\|\mathbb{M}\mathbf{1}_F\|_2^2}{\|\mathbf{1}_F\|_2^2} \geq 1 - |\mathbb{A}| \frac{|\partial F|}{|F|}. \tag{5.3.1}$$

If Γ is amenable, then by definition there are finite sets F for which the ratio on the right side of inequality (5.3.1) approaches 0, so it will follow directly from the lemma that $\|M\| = 1$.

Proof. Define the *inner boundary* of F to be the set $\partial_* F$ of all vertices in F that have at least one nearest neighbor not in F. Because the number of nearest neighbors of any vertex is the cardinality $|\mathbb{A}|$ of the generating set, the cardinalities of the inner

and outer boundaries are related by

$$|\mathbb{A}|^{-1}\,|\partial F| \leq |\partial_* F| \leq |\mathbb{A}|\,|\partial F|. \tag{5.3.2}$$

By definition, for any element $x \in F$ that is *not* in the inner boundary $\partial_* F$, the value of $\mathbf{1}_F$ is 1 at every nearest neighbor of x, and hence, since the random walk is assumed to be nearest neighbor, $\mathbb{M}\mathbf{1}_F(x) = 1$. Therefore,

$$\begin{aligned}
\|\mathbb{M}\mathbf{1}_F\|_2^2 &= \sum_{x\in\Gamma} |\mathbb{M}\mathbf{1}_F(x)|^2 \\
&\geq \sum_{x\in F\setminus\partial_* F} |\mathbb{M}\mathbf{1}_F(x)|^2 \\
&= |F| - |\partial_* F|,
\end{aligned}$$

and since $|F| = \|\mathbf{1}_F\|_2^2$, the inequality (5.3.1) follows. □

Exercise 5.3.3 Complete the proof of Kesten's theorem for amenable groups by establishing the following assertion: if Γ is amenable then for any symmetric random walk on Γ and any $\varepsilon > 0$ there exist finite sets $F \subset \Gamma$ such that

$$\frac{\|\mathbb{M}\mathbf{1}_F\|_2^2}{\|\mathbf{1}_F\|_2^2} \geq 1 - \varepsilon.$$

HINT: For any set $F \subset \Gamma$ and any integer $K \geq 1$ define $\partial_*^K F$ to be the set of points $x \in F$ at distance $\geq K$ from F^c. Show that for any $\varepsilon > 0$ there exist $K \in \mathbb{N}$ and finite sets $F \subset \Gamma$ such that

$$\begin{aligned}
|\partial_*^K F| < \varepsilon |F| \quad &\text{and} \\
\mathbb{M}\mathbf{1}_F(x) > 1 - \varepsilon \quad &\text{for all } x \in F \setminus \partial_*^K F.
\end{aligned}$$

5.4 The Dirichlet Form

The nonamenable case is not quite so simple, because to show that the Markov operator has norm $\|\mathbb{M}\| < 1$ one must effectively bound $\|\mathbb{M}f\|_2$ for *all* functions $f \in L^2(\Gamma)$, not just the indicators $f = \mathbf{1}_F$. For this purpose we introduce the *Dirichlet form.*

Definition 5.4.1 The *Dirichlet form* on $L^2(\Gamma)$ for a random walk with transition probabilities $p(\cdot,\cdot)$ is the bilinear function $\mathcal{D} : L^2(\Gamma) \times L^2(\Gamma) \to \mathbb{R}$ defined by

$$\mathcal{D}(f,g) = \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}(f(y)-f(x))(g(y)-g(x))p(x,y). \tag{5.4.1}$$

If the random walk is nearest-neighbor, then the only nonzero terms in the sum are those indexed by pairs (x, y) such that x and y are endpoints of an edge in the Cayley graph. Each edge is counted twice; this is the reason for including the factor $\frac{1}{2}$ in the definition (5.4.1).

Proposition 5.4.2 *If the random walk is* symmetric *then the Dirichlet form is well-defined and finite, and for any* $f \in L^2(\Gamma)$*,*

$$\mathcal{D}(f,f) = \langle f, (I-\mathbb{M})f\rangle . \tag{5.4.2}$$

Consequently,

$$1 - \|\mathbb{M}\| = \inf_{\|f\|_2=1} \mathcal{D}(f,f). \tag{5.4.3}$$

Proof. The equality $\|\mathbb{M}\| = \sup_{\|f\|_2=1}\langle f, \mathbb{M}f\rangle$ was established in Exercise 4.4.8, so (5.4.3) will follow directly from (5.4.2). To show that $\mathcal{D}(f,g)$ is finite it suffices, by the Cauchy-Schwarz inequality, to show that $\mathcal{D}(f,f) < \infty$ for any $f \in L^2(\Gamma)$. (Exercise: Check this.) That $\mathcal{D}(f,f) < \infty$ will follow from (5.4.2), because $|\langle f, \mathbb{M}f\rangle| \le \langle f, f\rangle$ for any f.

To prove (5.4.2), expand the square in the Dirichlet form and use the symmetry $p(x,y) = p(y,x)$ of the transition probabilities and the fact that $\sum_y p(x,y) = 1$ for all group elements x, y to obtain

$$\begin{aligned}
\mathcal{D}(f,f) &= \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}\Big(f(x)^2 + f(y)^2 - 2f(x)f(y)\Big)\,p(x,y) \\
&= \sum_x f(x)^2 - \sum_x\sum_y f(x)f(y)p(x,y) \\
&= \langle f, f\rangle - \langle f, \mathbb{M}f\rangle = \langle f, (I-\mathbb{M})f\rangle .
\end{aligned}$$

□

Exercise 5.4.3 Show that any function $f \in L^2(\Gamma)$ and all $n = 0, 1, 2, \cdots$,

$$\left\|\mathbb{M}^{2n+2}f\right\|_2 \le \left\|\mathbb{M}^{2n}f\right\|_2 \quad \text{and}$$
$$\left\|\mathbb{M}^{2n+2}f\right\|_2 - \left\|\mathbb{M}^{2n+4}f\right\|_2 \le \left\|\mathbb{M}^{2n}f\right\|_2 - \left\|\mathbb{M}^{2n+2}f\right\|_2 .$$

Notice that for the point mass $f = \delta_1$, the squared norms are the $2n$-step transition probabilities: $p_{2n}(1,1) = \|M^{2n}f\|_2^2$.

HINT: Use the Dirichlet form $\mathcal{D}_2$ associated with the two-step transition probabilities:

$$\mathcal{D}_2(f, f) := \frac{1}{2} \sum_{x \in \Gamma} \sum_{y \in \Gamma} (f(y) - f(x))^2 p_2(x, y). \tag{5.4.4}$$

5.5 Sobolev-Type Inequalities

By hypothesis (cf. Assumption 5.3.1), the step distribution of the random walk satisfies $\mu(y) > 0$ for every $y \in \mathbb{A}$. Consequently, there exists $\varepsilon > 0$ such that

$$p(1, y) = \mu(y) \geq \varepsilon \quad \text{for every} \quad y \in \mathbb{A}. \tag{5.5.1}$$

Definition 5.5.1 For any function $f : \Gamma \to \mathbb{R}$, define the *Sobolev seminorm* $\|f\|_S$ by

$$\|f\|_S := \frac{1}{2} \sum_{x \in \Gamma} \sum_{y \in \Gamma} |f(x) - f(y)| p(x, y). \tag{5.5.2}$$

Proposition 5.5.2 *If the transition probabilities are nearest-neighbor and satisfy* (5.5.1)*, then for every function* $f \in L^1(\Gamma)$*,*

$$\|f\|_S \geq \kappa \, \|f\|_1 \tag{5.5.3}$$

where $\kappa = \varepsilon\iota(G_{\Gamma;\mathbb{A}})/|\mathbb{A}|$ *and* $\|f\|_1 := \sum_{x \in \Gamma} |f(x)|$.

Proof. The inequality (5.5.1) implies that for every finite set $F \subset \Gamma$,

$$\sum_{x \in F} \sum_{y \notin F} p(x, y) \geq \varepsilon |\partial_* F| \geq \varepsilon\iota(G_{\Gamma;\mathbb{A}})/|\mathbb{A}|,$$

where $\partial_* F$ denotes the inner boundary of F, because for every $x \in \partial_* F$ there exists $y \notin F$ such that $p(x, y) \geq \varepsilon$.

Let $f \in L^1(\Gamma)$. Without loss of generality, we may assume that $f \geq 0$, since replacing f by $|f|$ can only decrease the Sobolev norm. For each $t > 0$, define F_t to be the set of all $y \in \Gamma$ for which $f(y) > t$. Clearly, if $f \in L^1(\Gamma)$ then each F_t is a finite set. Consequently,

$$\|f\|_S = \sum_{x, y \,:\, f(x) < f(y)} (f(y) - f(x)) p(x, y)$$

$$
\begin{aligned}
&= \int_0^\infty \sum_{x,y} \mathbf{1}\{f(x) < t < f(y)\}\, p(y,x)\, dt \\
&= \int_0^\infty \sum_{y \in F_t} \sum_{x \notin F_t} p(y,x)\, dt \\
&\geq \kappa \int_0^\infty |F_t|\, dt = \kappa \,\|f\|_1 .
\end{aligned}
$$

□

If Γ is amenable then the inequality (5.5.3) is vacuous. There is, however, a useful variant; this involves a function that we will call the *isoperimetric profile* of the Cayley graph.

Definition 5.5.3 The *isoperimetric profile* of the Cayley graph $G_{\Gamma;\mathbb{A}}$ is the nondecreasing function $J : \mathbb{N} \to \mathbb{R}_+$ defined by

$$J(n) := \max \left\{ \frac{|F|}{|\partial_* F|} \;:\; |F| \leq n \right\}. \tag{5.5.4}$$

Proposition 5.5.4 *Assume that the transition probabilities are nearest-neighbor and satisfy $p(1,a) \geq \varepsilon > 0$ for all $a \in \mathbb{A}$. Then for any finitely supported function $f : \Gamma \to \mathbb{R}$,*

$$\varepsilon \,\|f\|_1 \leq J(|support(f)|)\, \|f\|_S . \tag{5.5.5}$$

Exercise 5.5.5 Prove this.

5.6 Proof of Kesten's Theorem: Nonamenable Groups

To prove that $\|\mathbb{M}\| < 1$ it is enough, by Proposition 5.4.2, to show that there exists $\delta > 0$ such that $\langle f, (I - \mathbb{M} f)\rangle \geq \delta$ for every function $f \in L^2(\Gamma)$ with norm $\|f\|_2 = 1$, or equivalently,

$$\mathcal{D}(f,f) \geq \delta \quad \text{provided } \|f\|_2 = 1. \tag{5.6.1}$$

To accomplish this, we will use the Sobolev inequality to bound the L^2-norm of f by a multiple of the Sobolev norm, and then use the Cauchy-Schwarz inequality to bound the Sobolev norm by a multiple of the Dirichlet form $\mathcal{D}(f,f)$. First, with κ as in the Sobolev inequality (5.5.3),

$$\kappa \,\|f\|_2^2 = \kappa \left\| f^2 \right\|_1 \leq \left\| f^2 \right\|_S . \tag{5.6.2}$$

Next, by definitions of the Sobolev norm and Dirichlet forms,

$$\begin{aligned}
\left\| f^2 \right\|_S &= \frac{1}{2} \sum_x \sum_y |f(y)^2 - f(x)^2| p(x,y) \\
&= \frac{1}{2} \sum_x \sum_y |f(y) - f(x)||f(y) + f(x)| p(x,y) \\
&\le \left(\frac{1}{2} \sum_x \sum_y (f(y) - f(x))^2 p(x,y) \right)^{1/2} \\
&\quad \times \left(\frac{1}{2} \sum_x \sum_y (f(y) + f(x))^2 p(x,y) \right)^{1/2} \\
&\le \mathcal{D}(f,f)^{1/2} \left(\frac{1}{2} \sum_x \sum_y 4(f(y)^2 + f(x)^2) p(x,y) \right)^{1/2} \\
&= 2\mathcal{D}(f,f)^{1/2} \|f\|_2 \, . \qquad (5.6.3)
\end{aligned}$$

Squaring both sides shows that if $\|f\|_2 = 1$ then

$$\mathcal{D}(f,f) \ge (\kappa/2)^2,$$

which proves (5.6.1). □

Remark 5.6.1 This argument not only establishes that the spectral radius of the random walk is strictly less than 1, but provides the explicit upper bound

$$\varrho \le 1 - \left(\frac{\varepsilon \iota(G_{\Gamma;\mathbb{A}})|\mathbb{A}|^{-1}}{2} \right)^2 \quad \text{where} \quad \varepsilon = \min_{a \in \mathbb{A}} \mu(a). \qquad (5.6.4)$$

5.7 Nash-Type Inequalities

The key to Kesten's theorem in the nonamenable case was the Sobolev inequality (5.5.3), which in turn depended on the positivity of the isoperimetric constant (5.1.1). For amenable groups the "localized" Sobolev inequality (5.5.5) plays a similar role, but because the isoperimetric profile function J of an amenable Cayley graph grows unboundedly, the application of the inequality is more subtle.

Since the Sobolev inequality (5.5.5) involves the isoperimetric profile function J (cf. equation 5.5.4) in an essential way, it is imperative that we have good estimates

for it. The following proposition shows that J is controlled by the volume growth of the group.

Proposition 5.7.1 (Coulhon & Saloff-Coste [28]) *For any positive integer m define $\varrho(m) = \min\{k \geq 0 : |\mathbb{B}_k| \geq m\}$ to be the radius of the smallest ball in the Cayley graph that contains at least m points. Then for any finite set $F \subset \Gamma$,*

$$\frac{|F|}{|\partial_* F|} \leq 4\varrho(2|F|), \tag{5.7.1}$$

and so

$$J(m) \leq 4\varrho(2|F|). \tag{5.7.2}$$

Proof. For any element $a \in \mathbb{A}$ of the generating set $\mathbb{A}$ and any $x \in F$, the right translate $xa \notin F$ only if $x \in \partial_* F$. Consequently, for any element $g = a_1 a_2 \cdots a_k \in \Gamma$ of word length norm $|g| = k$, the number of points $x \in F$ such that $xg \notin F$ is at most $k|\partial_* F|$, because $xg \notin F$ only if one of the points

$$x, xa_1, xa_1a_2, \cdots, xa_1a_2 \cdots a_{k-1}$$

is an element of $\partial_* F$. Thus,

$$|\left\{(x, g) \in F \times \mathbb{B}_{\varrho(2|F|)} \; : \; xg \notin F\right\}| \leq |\mathbb{B}_{\varrho(2|F|)}|\varrho(2|F|)|\partial_* F|. \tag{5.7.3}$$

On the other hand, for any $x \in F$ at least half of the elements in the set $\left\{xg \; : \; g \in \mathbb{B}_{\varrho(2|F|)}\right\}$ are in F^c, because this set has cardinality $|\mathbb{B}_{\varrho(2|F|)}| \geq 2|F|$, by definition of the function $\varrho(\cdot)$. Therefore,

$$|\left\{(x, g) \in F \times \mathbb{B}_{\varrho(2|F|)} \; : \; xg \notin F\right\}| \geq \frac{1}{4}|F||\mathbb{B}_{\varrho(2|F|)}|. \tag{5.7.4}$$

The inequalities (5.7.3) and (5.7.4) imply (5.7.1), and this together with (5.3.2) implies (5.7.2). □

Corollary 5.7.2 *Suppose that the Cayley graph $G_{\Gamma;\mathbb{A}}$ has polynomial volume growth of degree $d \geq 1$, in the sense that for some constant $C > 0$*

$$|\mathbb{B}_m| \geq Cm^d \quad \textit{for all} \;\; m \in \mathbb{N}. \tag{5.7.5}$$

Then there exists $C' < \infty$ such that the isoperimetric profile $J(\cdot)$ of the Cayley graph satisfies

$$J(m) \leq C'm^{1/d}. \tag{5.7.6}$$

□

Exercise 5.7.3 Show that if the Cayley graph $G_{\Gamma;\mathbb{A}}$ satisfies the polynomial growth condition (5.7.5) for some finite, symmetric generating set $\mathbb{A}$ then it satisfies (5.7.5) (with a possibly different constant C, but the same d) for *every* finite, symmetric generating set.

Exercise 5.7.4 Show that if the isoperimetric profile $J(\cdot)$ satisfies (5.7.6) then the Cayley graph has polynomial volume growth of degree d, that is, show that (5.7.6) implies (5.7.5).

Proposition 5.7.5 (Nash Inequality) *Let $p(\cdot,\cdot)$ be the transition probabilities of a symmetric, nearest neighbor random walk on Γ, and assume that $p(x,y) \geq \varepsilon > 0$ for every pair $x, y \in \Gamma$ of neighboring elements. Let $\mathcal{D}$ be the associated Dirichlet form, and let J be the isoperimetric profile of the Cayley graph. Then for any function $f : \Gamma \to \mathbb{R}$ with finite support,*

$$\|f\|_2^2 \leq \frac{8}{\varepsilon^2} J\left(4\,\|f\|_1^2 \,/\, \|f\|_2^2\right)^2 \mathcal{D}(f,f). \tag{5.7.7}$$

Proof. As in the proof of Kesten's theorem (Section 5.6), we begin by using a Sobolev inequality — this time the local variant (5.5.5) — to bound the L^2-norm of f by a multiple of the Sobolev norm:

$$\varepsilon\,\|f\|_2^2 = \varepsilon \left\|f^2\right\|_1 \leq J(\mathrm{support}(f)) \left\|f^2\right\|_S. \tag{5.7.8}$$

Next, we recall the bound for the Sobolev seminorm $\|f^2\|_S$ established in (5.6.3):

$$\left\|f^2\right\|_S \leq 2\mathcal{D}(f,f)^{1/2}\,\|f\|_2\,.$$

This, together with (5.7.8), implies that for any compactly supported function f,

$$\varepsilon^2\,\|f\|_2^2 \leq 4J(\mathrm{support}(f))^2\mathcal{D}(f,f). \tag{5.7.9}$$

To obtain the inequality (5.7.7) from (5.7.9), we will employ a truncation technique that we learned from Pittet & Saloff-Coste [109]. Without loss of generality, we may assume that the function f is nonnegative, because changing f to $|f|$ can only decrease the right side of (5.7.7). For each $t \geq 0$, define F_t to be the set of all $x \in \Gamma$ where $f(x) > t$, and f_t to be the function $f_t(x) = \max(f(x), t) - t$. Since f is nonnegative, we have $F_t = \mathrm{support}(f_t)$, and so

$$t|F_t| = t|\{x : f(x) > t\}| \leq \|f\|_1\,. \tag{5.7.10}$$

Furthermore, for any $t \geq 0$,

$$f^2 \leq f_t^2 + 2tf \quad \text{and} \tag{5.7.11}$$

$$\mathcal{D}(f_t, f_t) \leq \mathcal{D}(f, f). \tag{5.7.12}$$

(Exercise: Check these!) Summing both sides of (5.7.11) (using the nonnegativity of f) and applying (5.7.9) to the function f_t yields

$$\begin{aligned} \|f\|_2^2 &\leq \|f_t\|_2^2 + 2t\,\|f\|_1 \\ &\leq \frac{4}{\varepsilon^2} J(|F_t|)^2 \mathcal{D}(f_t, f_t) + 2t\,\|f\|_1 \\ &\leq \frac{4}{\varepsilon^2} J(\|f\|_1 / t)^2 \mathcal{D}(f, f) + 2t\,\|f\|_1 . \end{aligned}$$

This holds for any $t \geq 0$; the trick now is to use a particular value of t that will turn the L^1-norm in the last term to an L^2-norm. Setting $4t = \|f\|_2^2 / \|f\|_1$ will accomplish this:

$$\|f\|_2^2 \leq \frac{4}{\varepsilon^2} J(4\,\|f\|_1^2 / \|f\|_2^2)^2 \mathcal{D}(f, f) + \frac{1}{2}\,\|f\|_2^2 ,$$

which is clearly equivalent to (5.7.9). □

Inequalities similar to (5.7.7) were first used by J. Nash [103] to prove theorems concerning the decay of heat kernels in Euclidean spaces. Their use in the study of random walks was apparently initiated by Varopoulos [127], who used them to establish (among other things) his celebrated growth criterion for recurrence, to which we shall return in Chapter 7. The basic idea behind Nash's program is crystallized in the following theorem, which in this form is due to T. Coulhon [27].

Theorem 5.7.6 (Nash; Coulhon) *Let $p(\cdot, \cdot)$ be the transition probabilities of a symmetric, nearest neighbor random walk on Γ, and assume that $p(x, y) \geq \varepsilon > 0$ for every pair $x, y \in \Gamma$ of neighboring elements. Assume that there is a positive, nondecreasing, continuous function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that for every compactly supported function $f : \Gamma \to \mathbb{R}$,*

$$\|f\|_2^2 \leq \Phi\left(\|f\|_1^2 / \|f\|_2^2\right)\left(\|f\|_2^2 - \|\mathbb{M}f\|_2^2\right), \tag{5.7.13}$$

where $\mathbb{M}$ is the Markov operator associated with the random walk. Then for all $n \in \mathbb{N}$,

$$p_{2n}(1, 1) \leq \Psi(n) \tag{5.7.14}$$

where $\Psi : \mathbb{R}_+ \to (0, 1]$ is the nonincreasing function defined by

$$t = \int_1^{1/\Psi(t)} \frac{\Phi(s)}{s}\, ds. \tag{5.7.15}$$

Remark 5.7.7 Let $\mathcal{D}_2$ be the Dirichlet form associated with the two-step transition probabilities $p_2(\cdot, \cdot)$; then by Proposition 5.4.2,

$$\begin{aligned}\mathcal{D}_2(f, f) := \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}(f(y) - f(x))^2 p_2(x, y) \\ = \left\langle f, (I - \mathbb{M}^2)f\right\rangle \\ = \|f\|_2^2 - \|\mathbb{M}f\|_2^2,\end{aligned}$$

the last because for a symmetric random walk the Markov operator $\mathbb{M}$ is self-adjoint (cf. equation (4.4.8)). This shows that the inequality (5.7.13) is essentially equivalent to the Nash inequality (5.7.7), with (a scalar multiple of) the isoperimetric profile J_2 for the Cayley graph $G_{\Gamma;\mathbb{A}^2}$ serving as the function Φ.

Exercise 5.7.8 Show that the function $\Psi : [0, \infty) \to (0, 1]$ implicitly defined by (5.7.15) is *strictly* decreasing, surjective (that is, $\Psi(0) = 1$ and $\lim_{t\to\infty}\Psi(t) = 0$), and satisfies the differential equation

$$\frac{d}{dt}\log\Psi(t) = \frac{\Psi'(t)}{\Psi(t)} = \frac{-1}{\Phi\left(1/\Psi(t)\right)}. \tag{5.7.16}$$

Proof of Theorem 5.7.6 For notational ease, let's write $u(n) = p_{2n}(1, 1)$. By Exercise 5.4.3, the function $u(n)$ is nonincreasing in n. The self-adjointness of the Markov operator implies that for every $n = 0, 1, 2, \cdots$,

$$\begin{aligned}u(n) = \left\langle\delta_1, \mathbb{M}^{2n}\delta_1\right\rangle = \left\|\mathbb{M}^n\delta_1\right\|_2^2 \quad \text{and} \\ u(n) - u(n+1) = \left\|\mathbb{M}^n\delta_1\right\|_2^2 - \left\|\mathbb{M}^{n+1}\delta_1\right\|_2^2,\end{aligned}$$

and because the random walk is symmetric,

$$\left\|\mathbb{M}^n\delta_1\right\|_1 = \sum_{x\in\Gamma} p_n(x, 1) = \sum_{x\in\Gamma} p_n(1, x) = 1.$$

Consequently, since the function Φ is nondecreasing, the inequality (5.7.13), with $f = \mathbb{M}^n\delta_1$, implies that

$$u(n) \leq \Phi(1/u(n))(u(n) - u(n+1)). \tag{5.7.17}$$

The remainder of the proof is an exercise in analysis, in which we will extract (5.7.14) from the inequality (5.7.17) and Exercise 5.7.8. Let $u(t)$ be the piecewise linear extension of u to $t \in \mathbb{R}_+$ gotten by linear interpolation between integers, that is, the unique continuous extension of u to $\mathbb{R}_+$ such that for each $n \in \mathbb{Z}_+$,

$$u'(t) = u(n+1) - u(n) \quad \text{for all} \;\; t \in (n, n+1). \tag{5.7.18}$$

This function is continuous and nonincreasing (in fact, strictly decreasing) for $t \in [0, \infty)$, and by definition $u(0) = 1 = \Psi(0)$. Furthermore, by (5.7.17),

$$u(t) \le u(n) \le -\Phi(1/u(n))u'(t) \quad \text{for all} \;\; t \in (n, n+1). \tag{5.7.19}$$

Now let $n \in \mathbb{Z}_+$ be any nonnegative integer such that $u(n) \le \Psi(n)$. By Exercise 5.7.8 the function Ψ is decreasing, and by hypothesis Φ is positive and nondecreasing; hence, the inequality (5.7.19) implies that for every $t \in [n, n+1]$

$$\frac{u'(t)}{u(t)} \le \frac{-1}{\Phi(1/u(n))} \le \frac{-1}{\Phi(1/\Psi(t))} = \frac{\Psi'(t)}{\Psi(t)}$$

for all $t \in (n, n+1)$. Integration shows that $u(t) \le \Psi(t)$ for all $t \in [n, n+1]$, and so the desired inequality (5.7.14) follows by induction. □

Corollary 5.7.9 (Varopoulos [126]) *Suppose that for some symmetric generating set $\mathbb{A}$ the Cayley graph $G_{\Gamma;\mathbb{A}}$ has polynomial volume growth of degree $d \ge 1$, in the sense that for some constant $C > 0$*

$$|\mathbb{B}_m| \ge Cm^d \quad \textit{for all} \;\; m \in \mathbb{N}. \tag{5.7.20}$$

Then for any irreducible, symmetric random walk on Γ whose step distribution has finite support there is a constant $C' \in (0, \infty)$ such that

$$p_n(1, x) \le C' n^{-d/2} \quad \textit{for all} \;\; n \ge 1 \;\; \textit{and all} \;\; x \in \Gamma. \tag{5.7.21}$$

Consequently, if $d > 2$ then every such random walk is transient.

In fact, neither symmetry nor finite support of the step distribution is needed for the validity of the estimate (5.7.21): see Lyons & Peres [91], Corollary 6.32. Observe that every infinite, finitely generated group Γ satisfies the growth condition (5.7.20) with $d = 1$, because the ball $\mathbb{B}_m$ contains at least one geodesic path of length m. Thus, Corollary 5.7.9 implies that the return probabilities of any symmetric, irreducible random walk on an infinite, finitely generated group whose step distribution has finite support must decay at least as fast as $p_n(1, 1) \le C/\sqrt{n}$.

Exercise 5.7.10 Show that if the step distribution of a symmetric random walk has support $\mathbb{A}$, where $\mathbb{A}$ is a finite, symmetric generating set, then its transition probabilities satisfy

$$p_{2n+1}(1, 1) \le |\mathbb{A}| p_{2n}(1, 1).$$

HINT: See Exercise 4.4.5.

Proof of Corollary 5.7.9 In view of Exercises 4.4.5 and (5.7.10), we need only establish the inequality (5.7.21) for $x = 1$ and for even integers $2n$. If μ is the step distribution of a symmetric, irreducible random walk on Γ then its support is a symmetric generating set, so without loss of generality we may assume that the generating set $\mathbb{A}$ has been chosen so that μ has support $\mathbb{A} \cup \{1\}$ and satisfies (5.5.1). By Theorem 5.7.6 and Proposition 5.7.5,

$$p_{2n}(1, 1) \leq \Psi(n) \quad \text{where} \quad t = \int_1^{1/\Psi(t)} \frac{C_1 J_2^2(4s)}{s} \, ds$$

for $C_1 = 8/\varepsilon^2$, where J_2 is the isoperimetric profile of the Cayley graph $G_{\Gamma;\mathbb{A}^2}$. (This choice of Cayley graph is dictated by the fact that the hypothesis (5.7.13) of Theorem 5.7.6 is equivalent to (5.7.7) for the two-step Dirichlet form $\mathcal{D}_2$; see Remark 5.7.7.) Corollary 5.7.2 implies that the isoperimetric profile J_2 satisfies $J_2(m) \leq C_2 m^{1/d}$, so for suitable constants $C_3, C_4 < \infty$,

$$\int_1^w \frac{C_1 J_2^2(4s)}{s} \, ds \leq C_3 w^{2/d} \quad \Longrightarrow \quad \Psi(t) \leq C_4 t^{-d/2}.$$

This proves (5.7.21). The final assertion, that $d > 2$ implies transience, follows directly from the estimate (5.7.21), by Pólya's Criterion. $\square$

Corollary 5.7.11 *Suppose that the Cayley graph $G_{\Gamma;\mathbb{A}}$ satisfies the growth condition (5.7.20) for some $d > 2$. Then for any symmetric, nearest-neighbor random walk there exists $C'' < \infty$ such that*

$$h(x) := P\{X_n = x \ \textit{for some} \ n \geq 0\} \leq C''|x|^{-(d-2)/2}. \tag{5.7.22}$$

Proof. Since the random walk is nearest-neighbor, $p_n(1, x) = 0$ for any $n < |x|$, and so

$$P\{X_n = x \ \text{for some} \ n \geq 0\} \leq \sum_{n \geq |x|} p_n(1, x).$$

$\square$

Inequality (5.7.22) forces an *upper* bound on the growth of balls in the Green metric d_G for a symmetric, nearest-neighbor random walk. Recall (cf Example 3.2.9) that d_G is the invariant metric on Γ defined by

$$d_G(x, y) = -\log h(x^{-1}y) \tag{5.7.23}$$

where h is the hitting probability function (5.7.22). Denote by $\mathbb{B}_r^G$ the ball of radius r for this metric, that is, $\mathbb{B}_r^G = \left\{x \in \Gamma \ : \ h(x) \geq e^{-r}\right\}$.

Proposition 5.7.12 *If the Cayley graph $G_{\Gamma;\mathbb{A}}$ satisfies the growth condition (5.7.20) for some $d > 2$, then for any symmetric, nearest-neighbor random walk there is a constant $C''' < \infty$ such that for every $r \geq 1$,*

$$|\mathbb{B}_r^G| \leq C''' \exp\{rd/(d-2)\}\,. \tag{5.7.24}$$

Proof. The *Green's function* of the random walk is defined by

$$G(x, y) := \sum_{n=0}^{\infty} p_n(x, y) = E^x \sum_{n=0}^{\infty} \mathbf{1}\{X_n = y\}\,; \tag{5.7.25}$$

thus, $G(x, y)$ is the expected number of visits to y by the random walk when started at x. By the Markov property, this expectation is the probability that the random walk ever visits the state y times the expected number of visits to y by the random walk when started at y, that is,

$$G(x, y) = h(x^{-1}y)G(1, 1) = \exp\{-d_G(x, y)\}\, G(1, 1).$$

Since $G(1, 1) \geq 1$, for any $x \in \mathbb{B}_r^G$ we must have $G(1, x) \geq \exp\{-r\}$.

Fix an integer m such that

$$\frac{1}{4}e^{-r} \leq \sum_{n=m}^{\infty} C' n^{-d/2} \leq \frac{1}{2}e^{-r},$$

where C' is the constant in Varopoulos' inequality (5.7.21). For suitable constants $0 < C_1 < C_2 < \infty$ we have

$$C_1 m \leq \exp\{dr/(d-2)\} \leq C_2 m.$$

Obviously, $\sum_{x\in\Gamma} p_n(1, x) = 1$ for any integer $n \geq 0$, and so

$$\begin{aligned}
m &= \sum_{x\in\Gamma} \sum_{n=0}^{m-1} p_n(1, x) \\
&\geq \sum_{x\in\mathbb{B}_r^G} \sum_{n=0}^{m-1} p_n(1, x) \\
&= \sum_{x\in\mathbb{B}_r^G} \left(\sum_{n=0}^{\infty} p_n(1, x) - \sum_{n=m+1}^{\infty} p_n(1, x) \right) \\
&\geq \frac{1}{2} |\mathbb{B}_r^G| \exp\{-r\}\,.
\end{aligned}$$

The inequality (5.7.24) now follows with $C''' = 2C_2$. □

Corollary 5.7.13 *The Green speed γ and the Avez entropy h of any symmetric, nearest-neighbor random walk on a finitely generated group Γ are equal.*

Recall that the Green speed γ is the limit $\gamma = \lim_{n\to\infty} n^{-1} E d_G(1, X_n)$ (cf. equation (3.2.5)). In fact, the equality $\gamma = h$ holds for any transient random walk whose step distribution has finite Shannon entropy: see Blachere, Haissinsky, and Mathieu [12].

Proof of Corollary 5.7.13 By Exercise 3.2.15, the inequality $\gamma \le h$ always holds, so it suffices to prove the reverse inequality $h \le \gamma$. Since both γ and h are nonnegative, we need only consider the case where $h > 0$. By Proposition 4.1.4, the exponential growth rate β of the group must be positive in order that a nearest-neighbor random walk have positive entropy. Consequently, we can assume that the group Γ has *super-polynomial* growth, that is, that (5.7.20) holds for *every* $d < \infty$.

Denote by β_G the exponential growth rate of Γ relative to the Green metric d_G, that is,

$$\beta_G = \lim_{m\to\infty} \frac{\log |\mathbb{B}_m^G|}{m}.$$

By Guivarc'h's "fundamental inequality" (cf. Exercise 3.2.14),

$$h \le \beta_G \gamma;$$

thus, to complete the proof we need only show that if (5.7.20) holds for every $d < \infty$ then $\beta_G \le 1$. But this follows directly from Proposition 5.7.12. □

Additional Notes. The dichotomy between amenable and nonamenable groups is of basic importance in modern group theory. See Paterson [106] for an extended account, including the role of nonamenability in the Banach-Tarski paradox and the Banach-Ruziewicz problem in measure theory.

Kesten's Theorem shows that the return probabilities of symmetric random walks on nonamenable groups decay exponentially, but does not provide sharp asymptotic estimates for these probabilities. Such estimates, for which the relative error of approximation converges to 0, are known as *local limit theorems*; the asymptotic formulas (1.5.3) for simple random walk on $\mathbb{Z}^d$ are the prototypical example. Local limit theorems for random walks on groups whose Cayley graphs are trees have been obtained in various degrees of generality by Woess [131], Gerl and Woess [49], and Lalley [84], [85]; see Woess [132] for a textbook presentation of some of these results. Cartwright [23] and Candellero and Gilch [20] prove similar results for other free products, showing among other things that "universality" need not hold in local limit theorems for symmetric, nearest-neighbor random walks on such groups. More recently, Gouëzel and Lalley [56] obtained sharp local limit theorems for symmetric random walks on cocompact Fuchsian groups (cf. Section 11.1.6 below for the

definition), and Gouëzel [54] subsequently extended this result to random walks on arbitrary hyperbolic groups.

Bounding the spectral radius of a symmetric random walk on a nonamenable group is an important research topic. A survey of results in this direction, with improvements and generalizations of the inequality (5.6.4), is given by Mohar and Woess [98]. Kesten proved in his articles [76, 77] that for random walk with the uniform distribution on a finite, symmetric generating set $\mathbb{A}$, the spectral radius ϱ satisfies the inequalities

$$\frac{\sqrt{|\mathbb{A}| - 1}}{|\mathbb{A}|} \leq \varrho \leq 1, \tag{5.7.26}$$

and that the lower bound is attained if and only if the Cayley graph is the infinite $|\mathbb{A}|$-ary tree. In the special case where Γ is the surface group of genus g, with the standard generating set, increasingly sharp bounds have been obtained by Zuk [135], Bartholdi et. al. [6], Nagnibeda [102], and Gouezel [55].

The presentation in Section 5.7 is adapted frrom the unpublished survey article [109] by Pittet and Saloff-Coste. There is a completely different approach, based on martingale methods, to the problem of deducing bounds on transition probabilities from the isoperimetric profile due to Morris and Peres [99]. A textbook exposition of this method can be found in Lyons and Peres [91], Chapter 6.

Chapter 6
Markov Chains and Harmonic Functions

6.1 Markov Chains

The study of transient random walks is inextricably tied to the theory of *harmonic functions* for the step distribution, as these functions determine *hitting probabilities* of various sorts. Much of the theory of harmonic functions does not require the homogeneity of the transition probabilities that comes with random walks; rather, the natural setting for the theory is the world of countable-state *Markov chains*. Random walks on finitely generated groups are Markov chains; as we will see, so are a number of auxiliary processes constructed from random walks.

Definition 6.1.1 A *Markov kernel* (also called a *transition probability matrix*) on a finite or countable set Θ (called the *state space*) is a function $p : \Theta \times \Theta \to [0, 1]$ such that

$$\sum_{y \in \Theta} p(x, y) = 1 \quad \text{for every } x \in \Theta. \tag{6.1.1}$$

A *Markov chain* with transition probability matrix p and initial point $x \in \Theta$ is a sequence of Θ-valued random variables $X_0, X_1, X_2, \cdots$ defined on a probability space[1] $(\Omega, \mathcal{F}, P^x)$ such that for each finite sequence $x_0, x_1, x_2, \cdots, x_m \in \Theta$,

$$P^x \left(\bigcap_{i=0}^{m} \{X_i = x_i\} \right) = \delta(x, x_0) \prod_{i=0}^{m-1} p(x_i, x_{i+1}), \tag{6.1.2}$$

[1] As for random walks, we indicate the initial point x of the Markov chain by a superscript on the probability measure. This is primarily for notational convenience. It can always be arranged that the probability measures P^x are all defined on the same measurable space $(\Omega, \mathbb{F})$.

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_6

where $\delta(\cdot,\cdot)$ is the Kronecker delta function. A Markov chain (or its transition probability matrix) is said to be *irreducible* if for any two states x, y there is a *positive probability path* from x to y, that is, a finite sequence $x = x_0, x_1, x_2, \cdots, x_m = y \in \Theta$ such that the probability (6.1.2) is positive. The *Markov operator* associated to a Markov kernel p is the linear transformation $\mathbb{M} : C_0(\Theta) \to C_0(\Theta)$ on the space $C_0(\Theta)$ finitely supported, real-valued functions u on Θ defined by

$$\mathbb{M}u(x) := \sum_{y\in\Theta} p(x, y)u(y) \quad \text{for all } x \in \Theta. \tag{6.1.3}$$

Exercise 6.1.2 (Chapman-Kolmogorov Equations) Let $(X_n)_{n\geq 0}$ be a Markov chain on Θ with transition probabilities $p(\cdot,\cdot)$, and for each $n \geq 1$ define the *n-step transition probabilities* $p_n(\cdot,\cdot)$ by

$$p_n(x, y) := P^x \{X_n = y\}. \tag{6.1.4}$$

Show that for any integers $n, m \geq 0$,

$$p_{n+m}(x, y) = \sum_{z\in\Theta} p_n(x, z)p_m(z, y). \tag{6.1.5}$$

Exercise 6.1.3 (Coset Markov Chain) Let Γ be a finitely generated group with a subgroup H, and let $(X_n)_{n\geq 0}$ be a random walk on Γ. Let $(Hx_i)_{i\in I}$ be an enumeration of the distinct right cosets of H in Γ, and let $\psi : \Gamma \to I$ be the function that assigns to each group element y the index i of its coset. Show that the sequence $(\psi(X_n))_{n\geq 0}$ is a Markov chain on the state space I, and show that if the random walk X_n is irreducible, then so is the Markov Chain $\psi(X_n)_{n\geq 0}$.

HINT: Use Proposition 6.1.4 below.

NOTE: If H is a *normal* subgroup then G/H is a group, and in this case the Coset Chain will be a random walk on the group G/H.

A transition probability matrix p on a state space Θ determines a *directed graph* $\mathbf{G} = (\Theta, \mathcal{E})$ with vertex set Θ and directed edge set

$$\mathcal{E} := \{(x, y) \in \Theta \times \Theta \; : \; p(x, y) > 0\}. \tag{6.1.6}$$

The digraph $\mathbf{G}$ is the natural analogue for a Markov chain of the Cayley graph $G_{\Gamma;\mathbb{A}}$ for a nearest-neighbor random walk on a finitely generated group Γ with generating set $\mathbb{A}$: in particular, a Markov chain $(X_n)_{n\geq 0}$ with transition probability matrix p is a (random) *path* in $\mathbf{G}$. Observe that $\mathbf{G}$ can have self-loops $(x, x) \in \mathcal{E}$, and vertices $x \in \Theta$ can have infinitely many *nearest neighbors* (points $y \in \Theta$ such that $(x, y) \in \mathcal{E}$).

For any Markov kernel p there exist probability spaces that support Markov chains with transition probability matrix p. One way to construct a Markov chain

uses an auxiliary sequence of independent, identically distributed random variables to provide random mappings of the state space to itself.

Proposition 6.1.4 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space equipped with a sequence of independent, identically distributed, random variables* $U_1, U_2, \cdots$ *taking values in a measurable space* $(\Upsilon, \mathcal{G})$*, and let* $F : \Theta \times \Upsilon \to \Theta$ *be a function such that for any two states* $x, y \in \Theta$*,*

$$P\{F(x, U_i) = y\} = p(x, y). \tag{6.1.7}$$

Then for any state x, the sequence $(X_n)_{n\geq 0}$ *of* Θ*-valued random variables defined inductively by*

$$X_0 = x \quad \text{and} \quad X_{n+1} = F(X_n, U_{n+1}) \tag{6.1.8}$$

is a Markov chain with transition probabilities $p(\cdot, \cdot)$ *and initial state* x*.*

Exercise 6.1.5 Prove this.

HINT: Verify (6.1.2) by induction, using Proposition A.3.4.

Exercise 6.1.6 Verify that for any Markov kernel p there is a Markov chain of the form (6.1.8) with i.i.d. random variables $U_1, U_2. \cdots$ that are uniformly distributed on the unit interval.

The construction (6.1.8) is called a *random mapping representation*. A random mapping representation is, in a certain sense, a (left) "random walk" on the semigroup of self-mappings of the state space Θ. This can be seen as follows. For each $n \in \mathbb{N}$, define $\Phi_n : \Theta \to \Theta$ to be the random mapping

$$\Phi_n(x) = F(x, U_n); \tag{6.1.9}$$

since the random variables U_n are independent and identically distributed, so are the random mappings $\Phi_1, \Phi_2, \cdots$. The inductive rule (6.1.8) can be rewritten as

$$X_n = \Phi_n \circ \Phi_{n-1} \circ \cdots \circ \Phi_1(X_0). \tag{6.1.10}$$

A Markov chain $\mathbf{X} = (X_0, X_1, X_2, \cdots)$ on a state space Θ can be viewed as a random variable taking values in the space Θ^∞ of one-sided sequences $\mathbf{x} = (x_n)_{n\geq 0}$ with entries in Θ. The relevant σ-algebra on Θ^∞ is the *Borel σ-algebra* $\mathcal{B}_\infty$, defined to be the σ-algebra generated by the coordinate random variables $\hat{X}_n : \Theta^\infty \to \Theta$ (that is, the minimal σ-algebra containing all events of the form $\left\{\hat{X}_n = x\right\}$).

Definition 6.1.7 The *law* (or *joint distribution*) of a Markov chain $\mathbf{X} = (X_0, X_1, X_2, \cdots)$ on the state space Θ with initial state x and transition probability kernel p is the induced probability measure

$$\hat{P}^x := P^x \circ \mathbf{X}^{-1}. \tag{6.1.11}$$

on $(\Theta^\infty, \mathcal{B}_\infty)$. Equivalently, $\hat{P}^x$ is the unique probability measure on $(\Theta^\infty, \mathcal{B}_\infty)$ such that for any finite sequence $x_0, x_1, \cdots, x_m \in \Theta$,

$$\hat{P}^x\left(\bigcap_{i=0}^{m}\left\{\hat{X}_i = x_i\right\}\right) = \delta(x, x_0)\prod_{i=0}^{m-1} p(x_i, x_{i+1}). \tag{6.1.12}$$

Clearly, under the measure $\hat{P}^x$ the sequence of coordinate random variables $\hat{X}_n$ is itself a Markov chain with the same initial point and transition probability matrix as $\mathbf{X} = (X_n)_{n\geq 0}$.

The defining property (6.1.2) has a trivial but important consequence, known as the *Markov property*. In its elemental form, this property asserts that probabilities of cylinder events factor. *Cylinder events* for Markov chains are defined as follows: for any sequence of states $x_0, x_1, \cdots$ and any integer $n \geq 0$,

$$C(x_0, x_1, \cdots, x_n) := \bigcap_{i=0}^{m}\{X_i = x_i\}. \tag{6.1.13}$$

Equation (6.1.2) implies that for any sequence $x_0, x_1, \cdots$ and all $m, n \geq 0$,

$$\begin{aligned} &P^{x_0}(C(x_0, x_1, \cdots, x_{m+n})) \\ &\qquad = P^{x_0}(C(x_0, x_1, \cdots, x_m))P^{x_m}(C(x_m, x_{m+1}, \cdots, x_{m+n})). \end{aligned} \tag{6.1.14}$$

This equation can be reformulated using the notation of *conditional probability* as follows: for any cylinder event $C = C(x_0, x_1, \cdots, x_m)$ with positive probability $P^{x_0}(C) > 0$, define the *conditional probability measure* $P^{x_0}(\cdot\,|\,C)$ on $\mathcal{F}$ by

$$P^{x_0}(F\,|\,C) = P^{x_0}(F \cap C)/P^{x_0}(C). \tag{6.1.15}$$

Proposition 6.1.8 (Markov Property) *If $C = C(x_0, x_1, \cdots, x_m)$ is a cylinder event with positive probability $P^{x_0}(C) > 0$ then the distribution of the sequence $(X_{n+m})_{n\geq 0}$ under the conditional measure $P^{x_0}(\cdot\,|\,C)$ is $\hat{P}^{x_m}$, that is, under $P^{x_0}(\cdot\,|\,C)$ the sequence $(X_{n+m})_{n\geq 0}$ is a Markov chain with transition probability matrix p and initial point x_m.*

Proof. This follows directly from (6.1.14), because measures on the σ-algebra $\sigma((X_n))_{n\geq m}$ are uniquely determined by their values on cylinder events, by Corollary A.2.3 of the Appendix. □

Definition 6.1.9 Let $(X_n)_{n\geq 0}$ be a Markov chain on a countable state space Θ. A state $z \in \Theta$ is *recurrent* if $P^z\{T_z < \infty\} = 1$, where $T_z = \min\{n \geq 1 : X_n = z\}$ is the time of first return to z. If the state z is not recurrent, then it is *transient*.

A Markov chain is said to be recurrent (respectively, transient) if every state is recurrent (respectively, transient).

Exercise 6.1.10 (a) Prove *Pólya's Criterion* for Markov chains: a state z is recurrent if and only if $\sum_{n=0}^{\infty} P^z\{X_n = z\} = \infty$. (b) Show that if the Markov chain X_n is *irreducible* then all states are of the same type (transient or recurrent).

We will mostly be interested in *transient* Markov chains. For transient chains, finite subsets of the state space are visited only finitely many times. It is natural to want to know where they go when they exit; for this purpose, we introduce the notion of a *first exit time*.

Definition 6.1.11 Let $(X_n)_{n\geq 0}$ be a Markov chain on a finite or countable state space Θ. For any nonempty subset $U \subset \Theta$, define τ_U, the *first exit time* from U, by

$$\tau_U = \min\{n \geq 0 \,:\, X_n \notin U\} \quad \text{or} \tag{6.1.16}$$

$$\tau_U = \infty \quad \text{if there is no such } n.$$

Lemma 6.1.12 *Let $(X_n)_{n\geq 0}$ be a Markov chain on Θ, and let $U \subset \Theta$ be a finite subset such that for every $x \in U$ there is a positive-probability path beginning at x and ending at a point $y \in U^c$. Then for every $x \in U$,*

$$P^x\{\tau_U < \infty\} = 1; \tag{6.1.17}$$

moreover, for some $\varrho \in (0, 1)$,

$$P^x\{\tau_U \geq n\} \leq \varrho^n \quad \textit{for all } \; n \in \mathbb{N}. \tag{6.1.18}$$

Proof. Since U is finite, there exists $m \in \mathbb{N}$ such that for every $x \in U$ there is a positive-probability path from x to U^c of length $\leq m$. Let $\varepsilon > 0$ be the minimum of the probabilities of these paths; then

$$\max_{z\in U} P^z\{\tau_U > m\} < 1 - \varepsilon.$$

This, together with the Markov property and a routine induction argument (exercise!), implies that for every $n \geq 1$,

$$\max_{z\in U} P^z\{\tau_U > nm\} < (1 - \varepsilon)^n.$$

□

The evolution of a Markov chain up to the time of first exit from a set $U \subset \Theta$ does not depend in any way on the transition probabilities $p(x, y)$ for states $x \notin U$, since the probabilities (6.1.2) of cylinder sets $C(x_0, x_1, \cdots, x_m)$ for which $x_i \in U$ for all $i < n$ involve only transition probabilities $p(x, y)$ for states $x \in U$. Thus,

changing the transition probabilities $p(x, y)$ for states $x \notin U$ will not change the behavior of the Markov chain up to the first exit time. One useful modification is as follows.

Definition 6.1.13 Let $p(\cdot, \cdot)$ be the transition probabilities of a Markov chain on Θ, and let $U \subset \Theta$ be a nonempty subset of Θ. The associated *Markov chain with absorption in* U^c is the Markov chain with transition probabilities

$$p^U(x, y) = p(x, y) \quad \text{if } \; x \in U, \tag{6.1.19}$$

$$p^U(x, y) = \delta(x, y) \quad \text{if } \; x \notin U. \tag{6.1.20}$$

Proposition 6.1.14 *Let $(X_n)_{n \geq 0}$ be a Markov chain on the state space Θ with transition probabilities $p(x, y)$, and let τ_U be the first exit time from a nonempty subset $U \subset \Theta$. Then the sequence $(X_{n \wedge \tau_U})_{n \geq 0}$ is a Markov chain with absorption in U^c, that is, it is a Markov chain on Θ with transition probabilities $p^U(x, y)$.*

Proof. Exercise. □

6.2 Harmonic and Superharmonic Functions

Definition 6.2.1 Let p be a Markov kernel on Θ. A function $h : \Theta \to \mathbb{R}$ is *harmonic* for the kernel p at a point $x \in U$ if the *mean value property*

$$h(x) = \sum_{y \in \Gamma} p(x, y) h(y) \tag{6.2.1}$$

holds. A function $h : \Theta \to \mathbb{R}$ is *superharmonic* (for the kernel p) at a point $x \in U$ if

$$h(x) \geq \sum_{y \in \Gamma} p(x, y) h(y). \tag{6.2.2}$$

For any subset $U \subset \Theta$ the function h is *harmonic (respectively, superharmonic) in U* if it is harmonic (superharmonic) at every $x \in U$, and it is *harmonic (superharmonic)* if it is harmonic (superharmonic) at every $x \in \Theta$.

Implicit in this definition is the assumption that the series in (6.2.1) and (6.2.2) are absolutely summable. If the function h is bounded, then this will always be the case. Clearly, linear combinations of harmonic functions are harmonic, and every constant function is harmonic.

Example 6.2.2 Fix a state $z \in \Theta$ and define $u_z : \Theta \to \mathbb{R}_+$ by $u_z(x) = P^x \left\{ \tau_{\Theta \setminus \{z\}} < \infty \right\}$, that is, $u_z(x)$ is the *hitting probability* of the state z. Then u_z is harmonic at every $x \neq z$ and superharmonic at $x = z$: see Exercise 6.2.10 below.

Example 6.2.3 † Recall that the free group $\mathbb{F}_2$ on two generators a, b is the set of all reduced words on the alphabet a, b, a^{-1}, b^{-1}; its Cayley graph (relative to this generating set) is the infinite homogeneous tree of degree 4. The simple random walk on $\mathbb{F}_2$ is the nearest-neighbor Markov chain with transition probabilities $p(x, y) = 1/4$ for all nearest-neighbor pairs $\{x, y\}$. For each positive real number θ set

$$f_\theta(a^{\pm 1}) = \frac{2e^{\pm\theta}}{1 + \cosh\theta} \quad \text{and} \quad f_\theta(b^{\pm 1}) = \frac{2}{1 + \cosh\theta}$$

and define $h_\theta : \mathbb{F}_2 \to \mathbb{R}_+$ by

$$h_\theta(w) = \prod_{i=1}^{m} f_\theta(x_i) \quad \text{where} \quad w = x_1, x_2 \cdots x_m$$

is the reduced word representation of w. It is easily checked (exercise!) that h_θ is harmonic. This example is of some interest because it provides an instance of a positive harmonic function h for which

$$\lim_{n\to\infty} h(X_n) = 0 \quad P^x - \text{almost surely} \tag{6.2.3}$$

for any $x \in \mathbb{F}_2$.

Exercise 6.2.4 † Prove this.

Exercise 6.2.5 Show that if h is a nonnegative harmonic (respectively, superharmonic) function then for every $n \in \mathbb{N}$ and every $x \in \Theta$,

$$h(x) \overset{(\geq)}{=} \sum_{y\in\Theta} p_n(x, y)h(y) \quad \text{where} \quad p_n(x, y) := P^x\{X_n = y\} \tag{6.2.4}$$

are the n-step transition probabilities. This implies the *Harnack inequalities*: if h is nonnegative and harmonic, then for any two states $x, y \in \Theta$ and any $n \in \mathbb{N}$,

$$h(x) \geq h(y)p_n(x, y). \tag{6.2.5}$$

Harmonic functions are of interest in Markov chain theory because they determine *hitting probabilities*. Conversely, Markov chains are of interest in potential theory because they provide integral representations of harmonic functions. The following propositions will make the connection clear.

Assumption 6.2.6 *Assume for the remainder of this section that $(X_n)_{n\geq 0}$ is a Markov chain on Θ with transition probability matrix p and that $U \subset \Theta$ is a nonempty, finite subset of the state space such that for every $x \in U$ there is a positive-probability path from x to U^c. Denote by ∂U the set of states $y \in U^c$ such*

that for some $x \in U$,

$$p(x, y) > 0.$$

Proposition 6.2.7 (Maximum Principle) *If $h : U \cup \partial U \to \mathbb{R}$ is harmonic in U then*

$$\max_{x \in U} h(x) \leq \sup_{x \in \partial U} h(x) \quad \textit{and}$$

$$\min_{x \in U} h(x) \geq \inf_{x \in \partial U} h(x).$$

Proof. Let $h_* = \max_{x \in U} h(x)$ (this is well-defined because U is finite), and let $W \subset U$ be the set of points where the maximum is attained in U. We claim that there is at least one $x \in W$ that has a nearest neighbor $y \in \partial U$ such that $p(x, y) > 0$ and $h(y) \geq h_*$.

By assumption, for every $x \in U$ there is a positive-probability path from x to ∂U. Every such path must exit the set W; consequently, there is at least one point $x \in W$ with a nearest neighbor $z \notin W$ such that $p(x, z) > 0$. Since h is harmonic at x,

$$\begin{aligned} h(x) = h_* &= \sum_{y \in W} p(x, y)h(y) + \sum_{y \in U \setminus W} p(x, y)h(y) + \sum_{y \in \partial U} p(x, y)h(y) \\ &= \sum_{y \in W} p(x, y)h_* + \sum_{y \in U \setminus W} p(x, y)h(y) + \sum_{y \in \partial U} p(x, y)h(y), \end{aligned}$$

and since x has at least one neighbor $z \notin W$ such that $p(x, z) > 0$, at least one of the terms in the second or third sum has positive weight $p(x, y)$. By definition of W, for any $y \in U \setminus W$ it must be the case that $h(y) < h_*$; therefore, there must be at least one $y \in \partial U$ such that $p(x, y) > 0$ and $h(y) \geq h_*$. □

Corollary 6.2.8 (Uniqueness Theorem) *Given a function $f : \partial U \to \mathbb{R}$ defined on the boundary of U, there is at most one function $h : U \cup \partial U \to \mathbb{R}$ that is harmonic in U and satisfies the boundary conditions*

$$h(x) = f(x) \quad \textit{for all} \;\; x \in \partial U.$$

Proof. The difference of two harmonic functions with the same boundary values would be a harmonic function in U with boundary values $\equiv 0$. The Maximum Principle implies that such a function must be identically 0 in U. □

Proposition 6.2.9 (Poisson Formula) *If $f : \partial U \to \mathbb{R}$ is* bounded, *then the function*

$$h(x) = E^x f(X_{\tau_U}) \tag{6.2.6}$$

is well-defined for $x \in U \cup \partial U$*, harmonic in* U*, and satisfies* $h = f$ *on* ∂U.

The equation (6.2.6) is a natural analogue of the *Poisson integral formula* of classical potential theory. It can be rewritten as

$$h(x) = \sum_{y \in \partial U} k(x, y) f(y) \quad \text{where} \quad k(x, y) = P^x \left\{ X_{\tau_U} = y \right\}; \tag{6.2.7}$$

the function $k : U \times \partial U \to [0, 1]$ is the *Poisson kernel* for the region U. Setting $f = \delta_y$ for some particular point $y \in \partial U$ shows that the hitting probability function $x \mapsto k(x, y)$ is itself harmonic in U, with boundary value function δ_y. Thus, if ∂U is finite the functions $k(\cdot, y)$ are a *basis* for the linear space of harmonic functions in U.

Proof of Proposition 6.2.9 If the starting point $X_0 = x$ of the Markov chain is an element of ∂U then the first exit time is $\tau_U = 0$, and so $E^x f(X_\tau) = E^x f(x) = f(x)$. Thus, $h = f$ on ∂U.

Suppose now that the starting point $X_0 = x$ is an element of U. By Lemma 6.1.12, the first exit time τ_U is finite, with probability one. Furthermore, since $x \in U$, the first exit time must satisfy $\tau_U \geq 1$, so we can compute the expectation $E^x f(X_\tau)$ by "conditioning on the first step":

$$E^x f(X_\tau) = \sum_{z \in \partial U} \sum_{y \in \Theta} f(z) P^x(\{X_1 = y\} \cap \{X_\tau = z\})$$

Thus, to show that the function $h(x) = E^x f(X_\tau)$ satisfies the mean value property in U it will suffice to show that

$$P^x(\{X_\tau = z\} \cap \{X_1 = y\}) = p(x, y) P^y \{X_\tau = z\}. \tag{6.2.8}$$

The probabilities on both sides of (6.2.8) can be calculated by summing over cylinder sets $C(x_0, x_1, \cdots, x_m) = \cap_{i=0}^m \{X_i = x_i\}$ for which $x_m = z$ and $x_i \in U$ for all $i \leq m - 1$. The cylinders that occur in the probability on the left side must all have $x_0 = x$ and $x_1 = y$, while those that occur on the right have $x_0 = y$. Consequently, there is a one-to-one correspondence between the cylinders C in the expectations on the left and right sides, given by

$$C(x, y, x_2, x_3, \cdots, x_m) \longleftrightarrow C(y, x_2, x_3, \cdots, x_m).$$

But the defining relation (6.1.2) for Markov chains implies that for any such matched pair of cylinders,

$$P^x(C(x, y, x_2, x_3, \cdots, x_m)) = p(x, y) P^y(C(y, x_2, x_3, \cdots, x_m));$$

summing over all pairs yields (6.2.8). □

Exercise 6.2.10 Show that the hitting probability function $u_z(x) = P^x \left\{\tau_{\Theta\setminus\{z\}} < \infty\right\}$ is harmonic at every $x \neq z$.

HINT: Let $\{x\} = U_0 \subset U_1 \subset \cdots$ be an increasing sequence of finite subsets of Θ whose union is $\Theta \setminus \{z\}$. Apply Proposition 6.2.9 to each of these.

Exercise 6.2.11 † Let $(X_n)_{n\geq 0}$ be the asymmetric nearest neighbor random walk on $\mathbb{Z}$ with transition probabilities $p(x, x+1) = r$ and $p(x, x-1) = 1-r$ where $0 < p \neq \frac{1}{2} < 1$. For integers $A, B \geq 1$ define $T = \tau_{(-A,B)} = \min\left\{n \geq 0 \,:\, S_{n \notin (A,B)}\right\}$. Find

$$P\{S_T = B\}.$$

HINT: Look for a harmonic function h such that $h(B) = 1$ and $h(A) = 0$. What happens when $r = \frac{1}{2}$?

Exercise 6.2.12 † Shows that if $h : \Theta \to \mathbb{R}_+$ is a nonnegative function that is superharmonic in the finite set U then for every $x \in U$,

$$h(x) \geq E^x h(X_{\tau_U}). \tag{6.2.9}$$

6.3 Space-Time Harmonic Functions†

If $(X_n)_{n\geq 0}$ is a Markov chain on a finite or countable set Θ, then the sequence $((X_n, n))_{n\geq 0}$ is also a Markov chain, albeit on an enlarged state space $\Theta \times \mathbb{Z}_+$. The transition probabilities q of this *space-time Markov chain* (X_n, n) are related to the transition probabilities p of X_n as follows:

$$\begin{aligned} q((x, n), (y, n+1)) &= p(x, y) \quad \text{and} \\ q((x, n), (y, m)) &= 0 \quad \text{if } m \neq n+1. \end{aligned} \tag{6.3.1}$$

Definition 6.3.1 A function $w : \Theta \times \mathbb{Z}_+ \to \mathbb{R}$ is *space-time harmonic* at (x, n) if it is harmonic for the space-time Markov chain, that is, if

$$w(x, n) = \sum_y p(x, y) w(y, n+1). \tag{6.3.2}$$

Proposition 6.3.2 *If $w : \Theta \times \mathbb{Z}_+ \to \mathbb{R}$ is space-time harmonic in a finite set $U' \subset \Theta \times \mathbb{Z}_+$, then for every $(x, m) \in U'$*

$$w(x, m) = E^{x,m} w(X_{\tau_{U'}}, \tau_{U'}) \quad \textit{where} \quad \tau_{U'} = \min\left\{n \geq m \,:\, (X_n, n) \notin U'\right\}. \tag{6.3.3}$$

Proof. This is a direct consequence of Proposition 6.2.9. (The hypothesis that there should exist a positive-probability exit path from any point of U' holds trivially, because since U' is finite there is a largest n such that U' intersects $\Theta \times [n, \infty)$.)

□

An important class of space-time harmonic functions is the set of renormalized eigenfunctions of the Markov operator $\mathbb{M}$ (which is defined by equation (4.4.2), with Θ in place of Γ). These are known as *α-harmonic functions*. More generally, if $f : \Theta \to \mathbb{R}$ satisfies the eigenvalue equation

$$\alpha f(x) = \sum_y p(x, y) f(y) \tag{6.3.4}$$

for all $x \in U$ then f is said to be *α-harmonic* in U. If f is α-harmonic, then the function $w(x, n) = \alpha^{-n} f(x)$ is space-time harmonic.

Corollary 6.3.3 *Assume that the Markov chain (X_n) satisfies the conditions of Assumption 6.2.6. If $f : \Theta \to \mathbb{R}_+$ is a nonnegative, bounded function that is α-harmonic for some $\alpha \geq 1$ in a finite set $U \subset \Theta$ then for every $x \in U$*

$$f(x) = E^x \alpha^{-\tau_U} f(X_{\tau_U}). \tag{6.3.5}$$

Proof. Proposition 6.3.2, with $U' = U \times [0, m]$, implies that for any $m \in \mathbb{Z}_+$ and any $x \in U$,

$$f(x) = E^x \alpha^{-\tau_U \wedge m} f(X_{\tau_U \wedge m}).$$

The corollary follows by the bounded convergence theorem. □

Example 6.3.4 † The function $h(x) = e^{\beta x}$ is $(\cosh \beta)^{-1}$-harmonic for the simple random walk $(X_n)_{n \geq 0}$ on $\mathbb{Z}$, so the identity (6.3.5) holds for every interval $U = [-A, 1]$. As $A \to \infty$, the first exit times τ_U converge monotonically up to $T = \min \{n \geq 0 : X_n = +1\}$. For *positive* values of β the function h is bounded on the negative halfline; consequently, the dominated convergence theorem can be applied in the $A \to \infty$ limit. This gives

$$E^0 e^{\beta} (\cosh \beta)^{-T} = 1.$$

Setting $s = 1/\cosh \beta$ and solving the resulting quadratic equation for $e^{-\beta}$ yields

$$e^{-\beta} = \frac{1 - \sqrt{1 - s^2}}{s}$$

leaving us with the generating function identity

$$E^0 s^T = \sum_{n=1}^{\infty} P^0 \{T = n\} s^n = \frac{1 - \sqrt{1 - s^2}}{s}.$$

Exercise 6.3.5 † Use Newton's binomial formula to deduce that

$$P^0 \{T = 2n - 1\} = (-1)^{n+1} \binom{1/2}{n}.$$

6.4 Reversible Markov Chains†

Definition 6.4.1 A transition probability matrix p on a state space Θ is *reversible* if there exists a positive function $\pi : \Theta \to \mathbb{R}$ (called a *total conductance measure*) such that for every pair of states $x, y \in \Theta$,

$$\pi(x) p(x, y) = \pi(y) p(y, x). \tag{6.4.1}$$

A transition probability matrix p is reversible relative to π if and only if the matrix $A = (a(x, y))_{x,y\in\Theta}$ with entries $a(x, y) := \pi(x)p(x, y)$ is symmetric. The entries $a(x, y) = a(y, x)$ can be interpreted as *electrical conductances* of wires connecting the vertices x, y. This interpretation provides a bridge between the theory of reversible Markov chains and electrical network theory that leads to some profound results, several of which will be developed (in the special case of Markov chains with *symmetric* transition probability matrices) in Chapter 7. For a more complete discussion, see the book [91].

Proposition 6.4.2 (Kolmogorov) *A transition probability matrix p on Θ is reversible if and only if for every* cycle $x = x_0, x_1, \cdots, x_m = x$ *of states*

$$\prod_{i=0}^{m-1} p(x_i, x_{i+1}) = \prod_{i=0}^{m-1} p(x_{i+1}, x_i). \tag{6.4.2}$$

Exercise 6.4.3 (A) Prove this. (B) Prove that if the transition probability matrix is irreducible and reversible then the total conductance measure π in (6.4.1) is unique up to scalar multiples.

Exercise 6.4.4 A transition probability matrix p on the infinite homogeneous tree $\Theta = \mathbb{T}_d$ is *nearest-neighbor* if $p(x, y) > 0$ only if the vertices $x, y \in \mathbb{T}_d$ are nearest neighbors. Show that any nearest-neighbor transition probability matrix on $\mathbb{T}_d$ is reversible.

Every *symmetric* transition probability matrix is reversible relative to the counting measure π on the state space. There are, however, reversible Markov chains — even reversible random walks on groups — whose transition probability matrices are

not symmetric. In fact, Exercise 6.4.4 implies that *every* nearest-neighbor random walk on the free group $\mathbb{F}_d$ is reversible. A special case is the asymmetric nearest-neighbor random walk on the integers $\mathbb{Z}$, that is, the random walk with step distribution

$$\mu(+1) = r = 1 - \mu(-1) \quad \text{where } r \in (0, 1). \tag{6.4.3}$$

Here the invariant measure π is the function

$$\pi(n) = \left(\frac{s}{r}\right)^n \quad \text{for } n \in \mathbb{Z}, \tag{6.4.4}$$

where $s = 1 - r$.

Proposition 6.4.5 *Let p be a reversible transition probability matrix on Θ with total conductance measure π. Define $L^2(\Theta, \pi)$ to be the Hilbert space of functions $f : \Theta \to \mathbb{R}$ that are square-summable with respect to π, that is, such that*

$$\begin{aligned} \|f\|_{2,\pi}^2 &:= \langle f, f\rangle_\pi < \infty \quad \textit{where} \\ \langle f, g\rangle_\pi &:= \sum_{x\in\Theta} f(x)g(x)\pi(x) \end{aligned} \tag{6.4.5}$$

is the inner product. Then the Markov operator $\mathbb{M}$ associated with p extends uniquely to a self-adjoint linear operator $\mathbb{M}$ on $L^2(\Theta, \pi)$ with operator norm $\|M\|_\pi \leq 1$, in particular, for every pair $f, g \in L^2(\Theta, \pi)$,

$$\langle f, \mathbb{M}g\rangle_\pi = \langle \mathbb{M}f, g\rangle_\pi \quad \textit{and} \tag{6.4.6}$$

$$\|Mf\|_{2,\pi} \leq \|f\|_{2,\pi}\,. \tag{6.4.7}$$

Proof. The Markov operator $\mathbb{M}$ is defined by the equation $\mathbb{M}f(x) := \sum_{y\in\Theta} p(x, y)f(y)$ for functions f with finite support. To prove that $\mathbb{M}$ extends to a linear operator on $L^2(\Theta, \pi)$ of norm ≤ 1, it suffices to show that $\|\mathbb{M}f\|_{2,\pi} \leq 1$ for every finitely supported function f such that $\|f\|_{2,\pi} = 1$. This is a simple exercise in the use of the Cauchy-Schwarz inequality and the reversibility hypothesis $\pi(x)p(x, y) = \pi(y)p(y, x)$:

$$\begin{aligned} \|\mathbb{M}f\|_{2,\pi} &= \sum_{x\in\Theta} \pi(x)(\mathbb{M}f(x))^2 \\ &= \sum_{x\in\Theta} \pi(x) \left(\sum_{y\in\Theta} p(x, y)f(y)\right)^2 \\ &\leq \sum_{x\in\Theta} \pi(x) \sum_{y\in\Theta} p(x, y)f(y)^2 \sum_{y\in\Theta} p(x, y) \end{aligned}$$

$$= \sum_{y\in\Theta}\sum_{x\in\Theta} \pi(y)p(y,x)f(y)^2 = \|f\|_{2,\pi}^2 .$$

The self-adjointness of $\mathbb{M}$ (equation (6.4.6)) follows by a routine calculation, using the definition (6.4.1) of reversibility and the fact that finitely supported functions are dense in $L^2(\Theta, \pi)$. For any two such functions f, g,

$$\begin{aligned}
\langle f, \mathbb{M}g\rangle_\pi &= \sum_{x\in\Theta} \pi(x)f(x)\mathbb{M}g(x) \\
&= \sum_{x\in\Theta} \pi(x)f(x) \sum_{y\in\Theta} p(x,y)g(y) \\
&= \sum_{y\in\Theta} \pi(y)g(y) \sum_{x\in\Theta} p(y,x)f(x) \\
&= \sum_{y\in\Theta} \pi(y)g(y)\mathbb{M}f(y) \\
&= \langle g, \mathbb{M}f\rangle_\pi .
\end{aligned}$$

□

Given Proposition 6.4.5, Carne-Varopoulos bounds for the transition probabilities of a reversible chain follow from Proposition 4.3.2 in much the same manner as for symmetric random walks (Section 4.4). Because the Kronecker delta functions δ_x have norms $\|\delta_x\|_{2,\pi}^2 = \pi(x)$ relative to the total conductance measure π, additional normalization factors appear in the bounds; the result is as follows.

Theorem 6.4.6 *If p is a reversible transition probability matrix on Θ with total conductance measure π, then for any two states $x, y \in \Theta$ and any integer $n \geq 1$,*

$$p_n(x,, y) \leq 2\sqrt{\frac{\pi(y)}{\pi(x)}}\, \|\mathbb{M}\|_\pi^n \exp\left\{-d(x,y)^2/2n\right\}. \tag{6.4.8}$$

Exercise 6.4.7 Check that the argument of Section 4.4 can be adapted to prove this.

Additional Notes. The *Poisson formula* (6.2.6) is the essential link between the theory of Markov processes and analysis. This formula dates at least as far back as the article of Courant, Friedrichs, and Lewy [29], who used it as the basis for a Monte Carlo method for the solution of the Dirichlet problem.

Harmonic functions figure prominently in the study of Markov chains and random walks in another way that we have not touched on. Suppose that $h : \Theta \to \mathbb{R}_+$ is a *positive* harmonic function for a Markov chain on Θ with transition probabilities $p(x, y)$; then the matrix $Q : \Theta \times \Theta \to \mathbb{R}$ with entries

$$q(x,y) := p(x,y)\frac{h(y)}{h(x)} \tag{6.4.9}$$

is itself a transition probability matrix. The transition probability matrix Q is known as Doob's *h-transform*. The h-transforms of the original transition probability matrix $P = (p(x, y))_{x,y\in\Theta}$ appear in the *ergodic decomposition* of P, at least when the Markov chain governed by P is transient. See Freedman [42], Chapter 4 for an account of this theory.

Chapter 7
Dirichlet's Principle and the Recurrence Type Theorem

If a finitely generated group Γ with symmetric generating set $\mathbb{A}$ supports a *recurrent*, irreducible, symmetric, nearest neighbor random walk, is it true that *every* irreducible, symmetric, nearest neighbor random walk is recurrent? Is it true that existence of a recurrent, symmetric, nearest-neighbor random walk implies that every irreducible, symmetric random walk whose step distribution has finite support (i.e., a symmetric random walk on another Cayley graph of the same group) is recurrent? The *Recurrence Type Theorem* answers both questions in the affirmative. In this chapter, we develop the machinery necessary — Dirichlet's Principle and Rayleigh's Comparison Principle — to prove this theorem.

7.1 Dirichlet's Principle

Symmetric Markov kernels on finite graphs — that is, those for which $p(x, y) = p(y, x)$ — are of interest in mathematical physics for reasons having nothing to do (at least directly) with probability theory. In particular, the values $p(x, y)$ can be viewed as the *electrical conductances* of metal wires connecting the endpoints x, y in an electrical network. In this context, harmonic functions with various boundary conditions are of interest because their values give the *voltages* at the nodes of the network when the boundary nodes are held at fixed values.

The correspondence between electrical networks and the theory of symmetric Markov chains leads to a host of interesting insights, many of which are explored in the book by Doyle and Snell [35] (or, in much greater depth, the book [91] by Lyons and Peres.). Here we will focus only on those aspects of the correspondence having to do with the *Dirichlet energy* (more properly, *power dissipation*) of the current flow associated with a voltage function.

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_7

Assumption 7.1.1 *Assume in Section 7.1 that p is a* ***symmetric*** *Markov kernel on a finite or countable space Θ with corresponding Markov operator $\mathbb{M}$, and let $(X_n)_{n\geq 0}$ be a Markov chain on Θ with transition probability matrix p.*

It will be useful for purposes of terminology and notation to view the state space Θ as the vertex set of the complete graph $G = (\Theta, \mathcal{E}_*)$, where $\mathcal{E}_*$ consists of all unordered pairs $\{x, y\}$ of distinct vertices. With this convention, the (outer) boundary ∂U of a nonempty set $U \subset \Theta$ coincides with the complement $U^c = \Theta \backslash U$.

Definition 7.1.2 For any set $U \subset \Theta$ of vertices, denote by $\mathcal{E}_U$ the set of edges (i.e., unordered pairs $\{x, y\}$ of distinct vertices) with at least one endpoint x or y in U. The *Dirichlet form* for the region U is the symmetric, nonnegative definite quadratic form

$$\mathcal{D}_U(g, g) = \mathcal{D}_U^p(g, g) = \sum_{\mathcal{E}_U} p(x, y)(g(y) - g(x))^2, \tag{7.1.1}$$

which is well-defined, but possibly infinite, for any function $g : \Theta \to \mathbb{R}$. The value $\mathcal{D}_U(g, g)$ of the Dirichlet form for a particular function g is the *Dirichlet energy* of the function in the region U. For $U = \Theta$ we will drop the subscript Θ: thus,

$$\mathcal{D} = \mathcal{D}_\Theta. \tag{7.1.2}$$

Note: If $x \in U$ then the sum (7.1.1) includes one term for every edge incident to x. For $y \notin U$, however, the sum only includes terms for those edges $\{x, y\}$ whose second endpoint x is in U. The sum in (7.1.1) can in general be infinite, but is finite for all bounded functions f if U is finite. For any function g that vanishes off U,

$$\mathcal{D}_U(g, g) = \mathcal{D}(g, g). \tag{7.1.3}$$

Proposition 7.1.3 (Dirichlet's Principle) *Let $U \subset \Theta$ be a finite set of vertices such that for every $x \in U$ there is a positive-probability path from x to U^c. Let $f : \partial U \to \mathbb{R}$ be a bounded real-valued function on the complement $\partial U = U^c$ of U. Then the unique minimizer of the Dirichlet energy $\mathcal{D}_U(h, h)$ among all functions $h : \Theta \to \mathbb{R}$ that coincide with f on $\Theta \setminus U$ is the harmonic function*

$$h(x) = E^x f(X_{\tau_U}). \tag{7.1.4}$$

Proof. The set $C(U; f)$ of functions $h : \Theta \to \mathbb{R}$ that agree with f outside U is a finite-dimensional vector space isomorphic to $\mathbb{R}^{|U|}$. The function $h \mapsto \mathcal{D}_U(h, h)$ is continuous on $C(U; f)$ (with respect to the Euclidean topology on $C(U; f) \cong \mathbb{R}^{|U|}$), and has compact sub-level sets, that is, for each $\alpha > 0$ the set of all $h \in C(U; f)$ such that $\mathcal{D}_U(h, h) \leq \alpha$ is compact. Consequently, the minimum value of $h \mapsto \mathcal{D}_U(h, h)$ is attained at some $h \in C(U; f)$.

Suppose that h is a minimizer of $\mathcal{D}_U(h, h)$. If there were some point $x \in U$ at which the mean value property failed, then the Dirichlet energy $\mathcal{D}_U(h, h)$ could be

reduced by changing the value of $h(x)$ to the mean value $\mathbb{M}h(x) = \sum_y p(x, y)h(y)$, because the contribution to $\mathcal{D}_U(h, h)$ from edges incident to x is

$$\begin{aligned}
&\sum_{y\in\Theta} p(x, y)(h(y) - h(x))^2 \\
&\qquad = \sum_{y\in\Theta} p(x, y)(h(y) - \mathbb{M}h(x) + \mathbb{M}h(x) - h(x))^2 \\
&\qquad = \sum_{y\in\Theta} p(x, y)(h(y) - \mathbb{M}h(x))^2 + \sum_{y\in\Theta} p(x, y)(h(x) - \mathbb{M}h(x))^2 \\
&\qquad = \sum_{y\in\Theta} p(x, y)(h(y) - \mathbb{M}h(x))^2 + (h(x) - \mathbb{M}h(x))^2.
\end{aligned}$$

Therefore, h must satisfy the mean value property at every $x \in U$. By Corollary 6.2.8 and Proposition 6.2.9, there is only one function with boundary value function f that satisfies the mean value property at each $x \in U$, and this satisfies equation (7.1.4). □

Exercise 7.1.4 † The *method of relaxations* is an iterative algorithm for numerically approximating the unique harmonic function h with boundary value f. Let $U = \{x_1, x_2, \cdots, x_m\}$ be an enumeration of the points of U. For any function $g \in C(U; f)$, define $R_i g \in C(U; f)$ to be the function obtained by replacing the value $g(x_i)$ by the mean value $\mathbb{M}g(x_i)$, that is,

$$\begin{aligned}
R_i g(x) &= g(x) \quad \text{for all} \quad x \neq x_i, \\
R_i g(x_i) &= \mathbb{M}g(x_i).
\end{aligned}$$

Now define $Rg = R_m \circ R_{m-1} \circ \cdots \circ R_1 g$ to be the function obtained by applying each R_i in sequence. Prove that, under the hypotheses of Proposition 7.1.3, for any $g \in C(U; f)$ the sequence $(R^n g)_{n\geq 1}$ converges pointwise to the unique harmonic function with boundary value f.

Since hitting probabilities are determined by harmonic functions (cf. Proposition 6.2.9), Dirichlet's Principle shows that they obey a *variational principle*. But there is more to the story than just this: in fact, the Dirichlet energy $\mathcal{D}_U(h, h)$ itself has a useful probabilistic interpretation. To derive this, we must introduce the notion of an *electrical current flow*.

Definition 7.1.5 A *flow* on a graph $G = (\Theta, \mathcal{E})$ is a function $J : \Theta \times \Theta \to \mathbb{R}$ such that

$$\begin{aligned}
J(x, y) &= -J(y, x) \quad \text{for all} \quad x, y \in \Theta, \\
J(x, y) &= 0 \quad \text{unless} \quad \{x, y\} \in \mathcal{E} \ \text{and}
\end{aligned}$$

$$\sum_{y\in\Theta} |J(x, y)| < \infty \quad \text{for every} \ \ x \in \Theta. \tag{7.1.5}$$

If J is a flow then $x \in \Theta$ is

(i) a *source* if $J(x+) := \sum_{y\in\Theta} J(x, y) > 0$;
(ii) a *sink* if $J(x+) < 0$; and
(iii) a *conservation point* if $J(x+) = 0$.

A flow on Θ is defined to be a flow on the complete graph $(\Theta, \mathcal{E}_*)$.

Example 7.1.6 If $h : U \cup \partial U \to \mathbb{R}$ is harmonic in a finite set $U \subset \Theta$ relative to the symmetric Markov kernel p then the *electrical current flow* defined by *Ohm's Law*

$$\begin{aligned} I(x, y) &= (h(x) - h(y))p(x, y) \quad \text{if} \ \ x \in U \ \ \text{or} \ \ y \in U, \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{7.1.6}$$

is conserved at all points $x \in U$.

Proposition 7.1.7 (Conservation Law) *Let $U \subset \Theta$ be a finite set of vertices in a graph $G = (\Theta, \mathcal{E})$ and let J a flow such that*

(a) $J(x, y) = 0$ unless $x \in U$ or $y \in U$; and
(b) J is conserved at every point of U.

Then for any bounded function $f : \Theta \to \mathbb{R}$ with finite support,

$$\sum_{\{x,y\}\in\mathcal{E}} (f(x) - f(y))J(x, y) = \sum_{x\in U^c} f(x)J(x+). \tag{7.1.7}$$

Note A: The sum on the left side of (7.1.7) is over all *undirected* edges $\{x, y\}$ of the graph. Since f has finite support the sum has only finitely many nonzero terms. Since any flow J is an antisymmetric function of the endpoints, the value of $(f(x) - f(y))J(x, y)$ does not depend on the order in which the endpoints x, y are listed.

Note B: Although we have dubbed this proposition a "conservation law", it is actually a discrete form of the divergence theorem of vector calculus.

Proof. The antisymmetry of J implies that for each edge $\{x, y\}$ we have $f(x)J(x, y) = -f(x)J(y, x)$; hence, because J is conserved in U,

$$\begin{aligned} \sum_{\{x,y\}\in\mathcal{E}} (f(x) - f(y))J(x, y) &= \sum_{x\in\Theta}\sum_{y\in\Theta} f(x)J(x, y) \\ &= \sum_{x\in U} f(x) \sum_{y\in\Theta} J(x, y) + \sum_{x\in U^c}\sum_{y\in\Theta} f(x)J(x, y) \end{aligned}$$

$$= 0 + \sum_{x \in U^c} \sum_{y \in U} f(x) J(x, y)$$
$$= 0 + \sum_{x \in U^c} f(x) J(x+).$$

□

Definition 7.1.8 For any vertex $z \in \Theta$ and any set $A \subset \Theta$ not containing z whose complement $U := \Theta \setminus A$ is finite, define the *escape probability* $p_{\text{escape}}(z; A)$ to be the probability that the Markov chain, when started in state $X_0 = z$, reaches A before returning to z, that is,

$$p_{\text{escape}}(z; A) = P^z \{T_A < T_z\} \quad \text{where} \quad T_B := \min \{n \geq 1 \; : \; X_n \in B\} . \tag{7.1.8}$$

Proposition 7.1.9 *The escape probability satisfies*

$$p_{escape}(z; A) = \min_h \mathcal{D}(h, h), \tag{7.1.9}$$

where the minimum is over all functions $h : \Theta \to \mathbb{R}$ that satisfy the boundary conditions

$$\begin{aligned} h(z) &= 1 \quad \textit{and} \\ h(y) &= 0 \quad \textit{for all } y \in A. \end{aligned} \tag{7.1.10}$$

Lemma 7.1.10 *Say that a set $V \subset \Theta$ is* accessible *from a point $y \notin V$ if there is a positive-probability path from y to a point $x \in V$. It suffices to prove Proposition 7.1.9 for sets A such that A is accessible from every point $y \in U = A^c$.*

Proof. Suppose first that A is not accessible from z; in this case, the escape probability $p_{\text{escape}}(z; A)$ is 0. Let $W \subset \Theta$ be the set of all vertices that are accessible from z; then $W \cap A = \emptyset$, and for any two vertices $w \in W$ and $v \notin W$ we must have $p(w, v) = 0$, because otherwise there would be a positive-probability path from z to v, and hence a positive-probability path from z to A. Consequently, if $h = \mathbf{1}_W$, then

$$\mathcal{D}(h, h) = 0,$$

and so formula (7.1.9) holds.

Suppose now that A is accessible from z. Let $V \subset U$ be the set of all vertices $y \in U$ such that A is *not* accessible from y. If $y \in V$, then y must be inaccessible from z, because otherwise, by the symmetry of the Markov chain, a positive-probability path from y to A could be constructed by concatenating a positive-probability path from y to z with a positive-probability path from z to A. Hence, moving y from U to A does not change the escape probability $p_{\text{escape}}(z; A)$. To prove the lemma,

then, we must show that replacing the set A by $A \cup V$ does not change the value of $\min_h \mathcal{D}(h, h)$.

Let $h : \Theta \to \mathbb{R}$ be any function that satisfies the boundary conditions (7.1.10), and let $h_* : \Theta \to \mathbb{R}$ be the function $h_* = h\mathbf{1}_{V^c}$, i.e,

$$\begin{aligned} h_*(y) &= h(y) && \text{for all } y \notin V, \\ h_*(v) &= 0 && \text{for all } y \in V. \end{aligned}$$

There are no positive-probability edges from either V to A or from V to $U \setminus V$, so

$$\begin{aligned} \mathcal{D}(h, h) &= \mathcal{D}_{V^c}(h, h) + \mathcal{D}_V(h, h) \\ &= \mathcal{D}(h_*, h_*) + \mathcal{D}_V(h, h) \\ &\geq \mathcal{D}(h_*, h_*). \end{aligned}$$

This proves that for every function h that satisfies the boundary conditions (7.1.10) there is a function h_* that satisfies the boundary conditions

$$\begin{aligned} h_*(z) &= 1 \quad \text{and} \\ h_*(y) &= 0 \quad \text{for all } y \in A \cup V \end{aligned}$$

such that $\mathcal{D}(h_*, h_*) \leq \mathcal{D}(h, h)$. Therefore, replacing A by $A \cup V$ does not change either side of the equation (7.1.9). □

Proof of Proposition 7.1.9. In view of Lemma 7.1.10, we may assume that the set U has the property that for every vertex $y \in U$ the set A is accessible from y. Thus, by Proposition 6.2.9, there is a unique harmonic extension φ to $U \setminus \{z\}$ of the boundary function (7.1.10), given by the Poisson formula (6.2.6):

$$\varphi(y) = P^y \left\{ \tau_{\{z\}} < \tau_A \right\} \quad \text{where} \;\; \tau_B := \min \{ n \geq 0 \,:\, X_n \in B \} .$$

Dirichlet's Principle (Proposition 7.1.3) implies that φ minimizes the energy $\mathcal{D}(h, h)$ among all functions satisfying the boundary conditions (7.1.10).

The escape probability can be written in terms of the function φ by the device of "conditioning on the first step", as in the proof of Proposition 6.2.9:

$$\begin{aligned} p_{\text{escape}}(z; A) &= \sum_{y \in \Theta} P^z(T_A < T_z \text{ and } X_1 = y) \\ &= \sum_{y \in \Theta} p(z, y) P^y(\tau_A < \tau_{\{z\}}) \\ &= \sum_{y \in \Theta} p(z, y)(1 - \varphi(y)) \end{aligned}$$

$$= \sum_{y \in U} p(z, y)(1 - \varphi(y)) + \sum_{a \in A} p(z, a).$$

(The change from T_A, T_z to $\tau_A, \tau_{\{z\}}$ is because of the time shift that results from changing the initial point of the Markov chain to y.)

Next, recall that $\mathcal{D}(h, h) = \mathcal{D}_U(h, h)$ for any function h that vanishes off U, and hence for any function that satisfies the boundary conditions (7.1.10). The set $\mathcal{E}_U$ that occurs in the sum (7.1.1) defining $\mathcal{D}_U$ consists of all edges of the complete graph G that have at least one endpoint in U. These can be partitioned into the set $\mathcal{E}_{U \setminus \{z\}}$ of edges that have at least one endpoint in $U \setminus \{z\}$, and the residual set $\mathcal{E}_{z,A} = \{\{z, a\} : a \in A\}$ connecting z to A. Thus,

$$\begin{aligned}
\mathcal{D}(\varphi, \varphi) = \mathcal{D}_U(\varphi, \varphi) &= \sum_{\mathcal{E}_U} p(x, y)(\varphi(x) - \varphi(y))^2 \\
&= \sum_{\mathcal{E}_{U \setminus \{z\}}} p(x, y)(\varphi(x) - \varphi(y))^2 + \sum_{a \in A} p(z, a)(1 - 0)^2 \\
&= \mathcal{D}_{U \setminus \{z\}}(\varphi, \varphi) + \sum_{a \in A} p(z, a).
\end{aligned}$$

Now let I be the electrical current flow defined by Ohm's law

$$\begin{aligned}
I(x, y) &= (\varphi(x) - \varphi(y))p(x, y) \quad \text{if} \quad \{x, y\} \in \mathcal{E}_{U \setminus \{z\}}, \\
&= 0 \quad \text{otherwise.}
\end{aligned}$$

Since φ is harmonic in $U \setminus \{z\}$, every point $x \in U \setminus \{z\}$ is a conservation point of I (cf. Example 7.1.6). Hence, by Proposition 7.1.7,

$$\begin{aligned}
\mathcal{D}_{U \setminus \{z\}}(\varphi, \varphi) &= \sum_{\mathcal{E}_{U \setminus \{z\}}} p(x, y)(\varphi(x) - \varphi(y))^2 \\
&= \sum_{\mathcal{E}_{U \setminus \{z\}}} I(x, y)(\varphi(x) - \varphi(y)) \\
&= \varphi(z) I(z+) = I(z+) \\
&= \sum_{y \in U} p(z, y)(1 - \varphi(y)).
\end{aligned}$$

Therefore,

$$\mathcal{D}_U(\varphi, \varphi) = \mathcal{D}_{U \setminus \{z\}}(\varphi, \varphi) + \sum_{a \in A} p(z, a)$$

$$= \sum_{y \in U} p(z, y)(1 - \varphi(y)) + \sum_{a \in A} p(z, a)$$

$$= p_{\text{escape}}(z; A).$$

□

7.2 Rayleigh's Comparison Principle

Proposition 7.2.1 (Comparison Principle) *Let p and q be two symmetric Markov kernels on a countable set Θ, with corresponding Dirichlet forms $\mathcal{D}^p$ and $\mathcal{D}^q$. Assume that there exists $C > 0$ such that for every bounded function $g : \Theta \to \mathbb{R}$ with finite support,*

$$C\mathcal{D}^p(g, g) \leq \mathcal{D}^q(g, g). \tag{7.2.1}$$

If the state $z \in \Theta$ is transient for the Markov chain with transition probability matrix p, then it is transient for the Markov chain with transition probability matrix q.

Proof. Let $\{z\} = U_0 \subset U_1 \subset U_2 \subset \cdots$ be a nested family of finite subsets of Θ that exhaust Θ, and set $A_n = \Theta \setminus U_n$. Proposition 7.1.9 implies that

$$p_{\text{escape}}(z, A_n) = \min \mathcal{D}^p(g, g) \quad \text{and}$$
$$q_{\text{escape}}(z, A_n) = \min \mathcal{D}^q(g, g),$$

where both minima are taken over the set of functions $g : \Theta \to \mathbb{R}$ that assume the value 1 at z and 0 on A_n. The hypothesis (7.2.1) implies that if the minimum values for the Markov kernel p are bounded away from 0, then so are the minimum values under q.

Since the sets A_n are decreasing in n, the escape probabilities $p_{\text{escape}}(z, A_n)$ (respectively, $q_{\text{escape}}(z, A_n)$) are nonincreasing, and therefore converge monotonically to the probability $P^z\{T_z = \infty\}$ (respectively, $Q^z\{T_z = \infty\}$) of no return. If p is transient, then $P^z\{T_z = \infty\} > 0$, so the escape probabilities $p_{\text{escape}}(z, A_n)$ are bounded away from 0. Thus,

$$P^z\{T_z = \infty\} > 0 \quad \text{implies} \quad Q^z\{T_z = \infty\} > 0.$$

□

This answers the first question posed at the beginning of this chapter: if a finite, symmetric generating set $\mathbb{A}$ supports a symmetric step distribution μ giving positive probability to every element of $\mathbb{A}$, and if the random walk with step distribution μ is recurrent, then *every* symmetric random walk with step distribution supported by

$\mathbb{A}$ is recurrent. The Recurrence Type Theorem, as stated in Section 1.4, states that recurrence of a single irreducible, symmetric, nearest-neighbor random walk on a finitely generated group Γ implies that *all* symmetric random walks with finitely supported step distributions are recurrent. The next theorem extends this assertion to all symmetric random walks whose step distributions have *finite second moment*, that is, any random walk $X_n = \xi_1\xi_2\cdots\xi_n$ whose step distribution μ satisfies

$$E|\xi_1|^2 = \sum_{x\in\Gamma} \mu(x)|x|^2 < \infty \tag{7.2.2}$$

where $|x|$ is the word-length norm of x. Obviously, the value $E|\xi_1|^2$ of the second moment depends on the choice of generating set, since this determines the word-length metric, but finiteness of $E|\xi_1|^2$ does not depend on the generating set (by inequalities (1.2.4)).

Theorem 7.2.2 (Recurrence Type Theorem) *Let Γ be an infinite, finitely generated group. If there is an* irreducible *recurrent symmetric random walk on Γ, then every symmetric random walk whose step distribution has finite second moment is recurrent. In this case the group Γ is called a* recurrent group.

Proof. Let $\mathbb{A}$ be a finite, symmetric generating set for Γ, and denote by q and $\mathcal{D}^q$ the transition probability kernel and Dirichlet form for the simple random walk on Γ (that is, the random walk whose step distribution is the uniform distribution on $\mathbb{A}$). To prove the theorem it will suffice to prove the following two assertions.

(A) If there exists a recurrent, irreducible, symmetric random walk then the simple random walk is also recurrent.
(B) If the simple random walk is recurrent, then so is every symmetric random walk whose step distribution has finite second moment.

Proof of (A). Let $(X_n)_{n\geq 0}$ be a recurrent, irreducible, symmetric random walk on Γ with transition probability kernel p. By Exercise 1.4.12, we may assume without loss of generality that this random walk is lazy. Since the random walk is both lazy and irreducible, there exists $m \in \mathbb{N}$ such that the m-step transition probability $p_m(1, x)$ is positive for every $x \in \mathbb{A}$, and without loss of generality we may assume that $m = 1$. (This is because if a lazy random walk $(X_n)_{n\geq 0}$ is recurrent, then so is the random walk $(X_{mn})_{n\geq 0}$, by Pólya's criterion; see Exercise 1.4.5.) Now if $p(1, x) > 0$ for every $x \in \mathbb{A}$ then the Dirichlet form $\mathcal{D}^p$ of the random walk dominates a multiple of the Dirichlet form $\mathcal{D}^q$ of the simple random walk. Since the random walk with transition probability kernel p is, by hypothesis, recurrent, it follows by Proposition 7.2.1 that the simple random walk is recurrent. □

Proof of (B). Let p and $\mathcal{D}^p$ be the transition probability kernel and Dirichlet form of a symmetric random walk $(X_n)_{n\geq 0}$ whose step distribution μ has finite second moment, and let q and $\mathcal{D}^q$ be the transition probability kernel and Dirichlet form of the simple random walk. We will show that the Dirichlet form $\mathcal{D}^p$ is dominated by a

constant multiple of the Dirichlet form $\mathcal{D}^q$. It will then follow by Proposition 7.2.1 that if the simple random walk is recurrent, then so is the random walk $(X_n)_{n\geq 0}$. $\square$

Let $x, y \in \Gamma$ be two points at (word) distance $d(x, y) = k$. By definition of the word metric, $y = xz$ for some $z = a_1 a_2 \cdots a_k$, where each $a_i \in \mathbb{A}$. Let $z_i = a_1 a_2 \cdots a_i$ be the ith point on the geodesic segment from 1 to $z = z_k$; then for any function $g : \Gamma \to \mathbb{R}$ with finite support,

$$|g(xz) - g(x)|^2 \leq \left(\sum_{i=1}^{k} |g(xz_i) - g(xz_{i-1})| \right)^2 \leq k \sum_{i=1}^{k} (g(xz_i) - g(xz_{i-1}))^2$$

the last by the Cauchy-Schwarz inequality. Summing over all $x \in \Gamma$, and using the fact that for each element $w \in \Gamma$ there are precisely k elements $x \in \Gamma$ such that $w = xz_i$ for some $i \leq k$, we obtain

$$\sum_{x\in\Gamma} |g(xz) - g(x)|^2 \leq k^2 \sum_{x\in\Gamma} \sum_{a\in\mathbb{A}} (g(xa) - g(x))^2 = 2k^2 |\mathbb{A}| \mathcal{D}^q(g, g).$$

It now follows that

$$\begin{aligned}
\mathcal{D}^p(g, g) &= \frac{1}{2} \sum_{x} \sum_{k=1}^{\infty} \sum_{|z|=k} p(x, xz)(g(xz) - g(x))^2 \\
&\leq \frac{1}{2} \sum_{k=1}^{\infty} \sum_{|z|=k} \mu(z)(2k^2 |\mathbb{A}| \mathcal{D}^q(g, g)) \\
&= \left\{ k^2 |\mathbb{A}| \sum_{z\in\Gamma} \mu(z)|z|^2 \right\} \mathcal{D}^q(g, g).
\end{aligned}$$

$\square$

Exercise 7.2.3 † It is not the case that every symmetric random walk on a recurrent group is recurrent. Take $\Gamma = \mathbb{Z}$: this is a recurrent group, by Pólya's theorem, but it admits transient symmetric random walks. Show, for instance, that the random walk with step distribution

$$\mu(2^n) = \mu(-2^n) = \frac{C}{2} \left(\frac{1}{\log_2 n} - \frac{1}{\log_2(n+1)} \right) \quad \text{for all } n \geq 3,$$

where $C = \log_2 3$, is transient.

HINT: Let $\xi_1, \xi_2, \cdots$ be i.i.d. with distribution μ. Let F_n be the event that $\max_{i\leq n} |\xi_i| \leq 2^n$ and G_n the event that the maximum is attained at more than one $i \leq n$. Show that $P(F_n \cap G_n \text{ i.o.}) = 0$. Then deduce from this that the random walk $S_n = \sum_{i=1}^{n} \xi_i$ returns to 0 only finitely many times.

Corollary 7.2.4 *A finitely generated group with a transient, finitely generated subgroup is transient.*

Proof. Suppose that the group Γ has a transient, finitely generated subgroup H. If Γ were recurrent, then by Theorem 7.2.2 every symmetric random walk on Γ with finitely supported step distribution would be recurrent. In particular, for any finitely generated subgroup H the simple random walk on H (relative to some finite, symmetric generating set) would be recurrent, and so H would be a recurrent group. □

7.3 Varopoulos' Growth Criterion

Varopoulos [129] used the Comparison Principle to give the following complete characterization of recurrent groups.

Theorem 7.3.1 (Varopoulos' Recurrence Criterion) *An infinite, finitely generated group is recurrent if and only if it has a finite-index subgroup isomorphic to either $\mathbb{Z}$ or $\mathbb{Z}^2$.*

The proof will require the following auxiliary result concerning the structure and growth of finitely generated groups.

Theorem 7.3.2 *An infinite, finitely generated group either has a finite-index subgroup isomorphic to $\mathbb{Z}$ or $\mathbb{Z}^2$ or it has at least cubic growth, that is, there exists a constant $\kappa > 0$, such that for every m the ball $\mathbb{B}_m$ of radius m in the Cayley graph has volume*

$$|\mathbb{B}_m| \geq \kappa m^3. \tag{7.3.1}$$

This is a consequence of a deep theorem of Gromov [60], according to which every finitely generated group of polynomial growth is *virtually nilpotent*, together with a theorem of Bass [7] and (independently) Guivarc'h [62] regarding the growth of nilpotent finitely generated groups. A somewhat simpler proof of Theorem 7.3.2 that does not rely on any of these theorems will be given in Section 15.4 below.

In Section 7.4, we will prove the *Rigidity Theorem* for recurrence type stated in Chapter 1, according to which the recurrence type of a finitely generated group is the same as that of any finite-index subgroup. Since both $\mathbb{Z}$ and $\mathbb{Z}^2$ are recurrent, by Pólya's Theorem (cf. Section 1.5), the *if* implication of Theorem 7.3.1 will follow. In view of Theorem 7.3.2, the converse implication will follow from the following proposition, also due to Varopoulos.

Proposition 7.3.3 *Any finitely generated group with at least cubic growth is transient.*

Proof. By Theorem 7.2.2, it is enough to exhibit a transient, symmetric random walk on the group whose step distribution has finite second moment. But this has

already been done in Corollary 5.7.9, which implies that if the growth condition (7.3.1) holds then the simple nearest neighbor random walk is transient. □

7.4 Induced Random Walks on Subgroups

Theorem 7.4.1 *A finite-index subgroup of a finitely generated group Γ has the same recurrence type as Γ.*

This is the *Rigidity Theorem* for recurrence type whose proof was promised in Section 1.4. Like Varopoulos' Theorem, the Rigidity Theorem can be proved by means of the Comparison Principle. We will take a different tack here, however, and instead deduce the theorem from an analysis of the *induced random walk* on the subgroup, which we construct in Proposition 7.4.3 below.

Corollary 7.2.4 implies that if a finitely generated group Γ has a transient, finitely generated subgroup then Γ is also transient. Since finite-index subgroups are finitely generated (Exercise 1.2.11), it follows that to prove Theorem 7.4.1 it will suffice to prove that if Γ has a recurrent, finite-index subgroup then Γ is recurrent.

Assumption 7.4.2 *Assume for the remainder of this section that Γ is a finitely generated group with finite-index subgroup H, and assume that $(X_n)_{n\geq 0}$ is an irreducible random walk on Γ with step distribution μ.*

Proposition 7.4.3 *Define nonnegative integer-valued random variables $T_0, T_1, \cdots$ inductively by*

$$T_0 = 0 \quad \text{and} \quad T_{m+1} = \min\{n \geq T_m + 1 \; : \; X_n \in H\}. \tag{7.4.1}$$

Then

(A) $P^x\{T_m < \infty\} = 1$ for every $m \in \mathbb{N}$ and every $x \in \Gamma$.
(B) If $x \in H$ then the sequence $(X_{T_m})_{m\geq 0}$ is a random walk under P^x.
(C) The step distribution of this random walk does not depend on the initial point $x \in H$.
(D) If the random walk $(X_n)_{n\geq 0}$ is symmetric and irreducible then so is $(X_{T_m})_{m\geq 0}$.
(E) If μ has finite support then $E|X_{T_1}|^2 < \infty$.

We will call the sequence $(X_{T_m})_{m\geq 0}$ the *induced random walk* on H. For brevity, write $T = T_1$. Before proving Proposition 7.4.3, let's see how it implies Theorem 7.4.1.

Proof of Theorem 7.4.1. As we have already observed, it suffices to prove that if H is recurrent then so is Γ. For this it suffices, by Theorem 7.2.2, to show that for some finite, symmetric generating set $\mathbb{A}$ of Γ the simple random walk on Γ (step distribution = uniform distribution on $\mathbb{A}$) is recurrent.

Let $(X_n)_{n\geq 0}$ be the simple random walk on Γ, and let $(X_{T_m})_{m\geq 0}$ be the induced random walk on H. By Proposition 7.4.3, the induced random walk is symmetric

and irreducible, and its step distribution has finite second moment. Therefore, since the subgroup H is by hypothesis recurrent, the induced random walk is recurrent. Consequently, by Corollary 1.4.6, the sequence $(X_{T_m})_{m\geq 0}$ makes infinitely many returns to the group identity 1, with probability one. It follows that the simple random walk $(X_n)_{n\geq 0}$ also makes infinitely many returns to 1, and this implies that it is recurrent. □

Proposition 7.4.3 also leads to an easy proof of the amenability criterion established in Exercise 5.1.11.

Corollary 7.4.4 *If Γ has an amenable subgroup H of finite index then Γ is amenable.*

Proof. Let $(X_n)_{n\geq 0}$ be the simple, nearest-neighbor random walk on Γ with respect to a finite, symmetric generating set $\mathbb{A}$. By Proposition 7.4.3, the induced random walk $(X_{T_m})_{m\geq 0}$ on H is symmetric, so by Kesten's Theorem, if H is amenable then

$$\lim_{m\to\infty} P\left\{X_{T_{2m}} = 1\right\}^{1/2m} = 1 \quad \Longrightarrow$$

$$\sum_{m=1}^{\infty} R^{2m} P\left\{X_{T_{2m}} = 1\right\} = \infty \quad \text{for all } R > 1.$$

Since $T_{2m} \geq 2m$, and since $(X_{T_m})_{m\geq 0}$ is a subsequence of $(X_n)_{n\geq 0}$, it follows (Exercise: explain why!) that

$$\sum_{n=1}^{\infty} R^{n} P\left\{X_{n} = 1\right\} = \infty \quad \text{for all } R > 1.$$

Consequently, the simple random walk $(X_n)_{n\geq 0}$ on Γ has spectral radius 1, and so Kesten's Theorem implies that Γ is amenable. □

The remainder of this section will be devoted to the proof of Proposition 7.4.3. We will break the argument into four steps.

Lemma 7.4.5 *There exist constants $C < \infty$ and $0 < \varrho < 1$ such that for every $x \in \Gamma$*

$$P^x\left\{T_1 \geq n\right\} \leq C\varrho^n \quad \textit{for all } \; n \in \mathbb{N}. \tag{7.4.2}$$

Consequently, for every $x \in \Gamma$,

$$E^x T_1 \leq C\varrho/(1-\varrho), \tag{7.4.3}$$

and so for each $x \in H$ the random variables T_m are P^x-almost surely finite.

Proof. Let $(Hx_i)_{i\leq K}$ be an enumeration of the right cosets of H, and let ψ : $\Gamma \to [K]$ be the mapping that assigns to each group element the index of its

coset. In Exercise 6.1.3, you verified that the sequence $(\psi(X_n))_{n\geq 0}$ is an irreducible Markov chain on the finite state space $[K]$. Consequently, the assertion follows from Lemma 6.1.12, which implies that for any irreducible, finite-state Markov chain the first-passage time to any point has a distribution with geometrically decaying tail. The bound (7.4.3) follows immediately from (7.4.2). The final statement follows by the Markov property and induction, which imply that for $m, n \in \mathbb{Z}_+$ and each $x \in \Gamma$,

$$P^x\{T_{m+1} - T_m \geq n\} = \sum_{k=0}^{\infty} \sum_{y\in H} P^x\{T_m = k \text{ and } X_k = y\} P^y\{T_1 \geq n\} \leq C\varrho^n.$$

□

Lemma 7.4.6 *For each $m = 1, 2, \cdots$ let*

$$Y_m = X_{T_{m-1}}^{-1} X_{T_m} \tag{7.4.4}$$

be the mth step of the sequence $(X_{T_m})_{m\geq 0}$. Then the random vectors $(Y_m, T_m - T_{m-1})_{m\geq 1}$ are independent and identically distributed under P^x, with the same joint distribution as under P^1. Consequently, the sequence $(X_{Y_m})_{m\geq 0}$ is, under P^x, a random walk on H with initial point x.

Proof. Assume first that the initial point $x = 1$ is the group identity. We will proceed by induction on m. Lemma 7.4.5 implies that $P\{T_1 < \infty\} = 1$, and so the random variable Y_1 is well-defined. Denote by $\xi_n = X_{n-1}^{-1} X_n$ the steps of the random walk X_n, and for any sequence $(z_n)_{n\in\mathbb{N}}$ in Γ and any $n \in \mathbb{N}$ define the cylinder events

$$C(z_1, z_2, \cdots, z_n) = \bigcap_{i=1}^{n} \{\xi_i = z_i\}.$$

Because the random variables ξ_n are independent and identically distributed, the probabilities of these cylinder events factor: for any integers $n, k \geq 0$,

$$P(C(z_1, z_2, \cdots, z_{n+k})) = P(C(z_1, z_2, \cdots, z_n))P(C(z_{n+1}, z_{n+2}, \cdots, z_{n+k})).$$

Let $C(z_1, z_2, \cdots, z_n)$ be a cylinder on which $T_m = n$. Fix $y \in H$ and $k \in \mathbb{N}$, and sum over all $(z_{n+1}, z_{n+2}, \cdots, z_{n+k})$ such that

$$k = \min\left\{j : z_{n+1}z_{n+2}\cdots z_{n+j} \in H\right\} \quad \text{and} \quad z_{n+1}z_{n+2}\cdots z_{n+k} = y;$$

this gives

$$P(C(z_1, z_2, \cdots, z_n) \cap \{T_{m+1} - T_m = k\} \cap \{Y_{m+1} = y\}) \\ = P(C(z_1, z_2, \cdots, z_n))P(\{T_1 = k\} \cap \{Y_1 = y\}).$$

Since $P\{T_1 < \infty\} = 1$, it follows by induction (by summing over all $k \in \mathbb{N}$, $y \in H$, and all cylinders on which $T_m < \infty$) that $P\{T_{m+1} < \infty\} = 1$. Similarly, summing over all cylinders $C(z_1, z_2, \cdots, z_n)$ on which

$$T_j - T_{j-1} = n_j \quad \text{and} \quad Y_j = y_j \text{ for all } j = 1, 2, \cdots, m$$

shows that

$$\begin{aligned} &P\{T_j - T_{j-1} = n_j \text{ and } Y_j = y_j \text{ for all } j \le m+1\} \\ &\quad = P\left\{T_j - T_{j-1} = n_j \text{ and } Y_j = y_j \text{ for all } j \le m\right\} \\ &\qquad \times P\{T_1 = n_{m+1} \text{ and } Y_1 = y_{m+1}\} \\ &\quad = \prod_{j=1}^{m+1} P\left\{T_1 = n_j \text{ and } Y_1 = y_j\right\}. \end{aligned}$$

This proves that the random vectors $(T_{m+1} - T_m, Y_m)$ are independent and identically distributed under $P = P^1$. The case where the initial point x is arbitrary follows by a similar argument. □

Lemma 7.4.7 *If $(X_n)_{n\ge0}$ is a symmetric, irreducible random walk on Γ then the induced random walk $(X_{T_m})_{m\ge0}$ on the subgroup H is also symmetric and irreducible.*

Exercise 7.4.8 Prove this.

HINT: To prove symmetry, reverse and invert the steps of the random walk between successive visits to H.

Lemma 7.4.9 *If $(X_n)_{n\ge0}$ is an irreducible random walk on Γ whose step distribution μ has finite support, then the step distribution of the induced random walk on the subgroup H has finite second moment.*

Proof. This is an easy consequence of Lemma 7.4.5, because if μ has finite support, then $|X_{T_1}| \le CT_1$ for some constant $C < \infty$ depending only on the support. Lemma 7.4.5 implies that the random variable T_1 has all moments finite, so in fact $E|X_{T_1}|^k < \infty$ for every $k \in \mathbb{N}$. □

Proof of Proposition 7.4.3. Lemma 7.4.5 implies (A); Lemma 7.4.6 implies (B) and (C); Lemma 7.4.7 implies (D); and Lemma 7.4.9 implies (E). □

Additional Notes. The use of electrical network theory to obtain criteria for the recurrence of a symmetric Markov chain begins with the seminal work [104] of Nash-Williams. His criterion is discussed in the book [35] of Doyle and Snell,

where it is used to give an interesting proof of the recurrence of simple random walk on $\mathbb{Z}^2$. Another noteworthy such result is the "finite energy flow" criterion of T. Lyons [93]. The "Rigidity Theorem" of Section 7.4 is usually proved by using Rayleigh's comparison method: see, for instance, Woess [132], Chapter 1. However, our proof, based on the properties of the induced random walk enumerated in Proposition 7.4.3, seems interesting in its own right.

Chapter 8
Martingales

Together with the Uniqueness Theorem (Corollary 6.2.8), the Poisson formula (6.2.6) implies that for any irreducible Markov chain, every harmonic function in a finite region U is uniquely determined by its values on the boundary ∂U. Consequently, for any finite region U the vector space of real harmonic functions in U is isomorphic to the space of real-valued functions on the boundary ∂U. Is there an analogous characterization of the space of harmonic functions on the entire state space? The answer, as we will show in Chapter 9, is tied up with the limiting behavior of harmonic functions along Markov chain paths.

Theorem 8.0.1 *If $(X_n)_{n\geq 0}$ is a Markov chain on Θ and $h : \Theta \to \mathbb{R}_+$ is nonnegative and harmonic everywhere in Θ, then for any initial point $x \in \Theta$,*

$$P^x \left\{ \lim_{n\to\infty} h(X_n) := W \in \mathbb{R} \;\; \textit{exists} \right\} = 1. \tag{8.0.1}$$

Furthermore, if the function h is bounded then for every $x \in \Theta$,

$$h(x) = E^x W. \tag{8.0.2}$$

Exercise 8.0.2 Deduce from Theorem 8.0.1 that an irreducible, *recurrent* Markov chain admits no nonconstant nonnegative harmonic functions.

The integral formula (8.0.2) is a natural analogue of the Poisson formula (6.2.6). But whereas (6.2.6) provides a complete description of the vector space of harmonic functions in a finite region U, the equation (8.0.2) does not by itself give an analogous description of the vector space of all bounded harmonic functions, because it does not provide a characterization of the space of possible limit random variables W. In Section 9.1, we will complete the story by showing that the possible limits W are precisely the (bounded) *invariant random variables*. In this chapter we will show how to deduce Theorem 8.0.1 from *martingale inequalities*.

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_8

8.1 Martingales: Definitions and Examples

Definition 8.1.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ be a nested sequence of σ-algebras all contained in $\mathcal{F}$, henceforth called a *filtration* of $(\Omega, \mathcal{F}, P)$. A sequence of random variables $M_0, M_1, M_2, \cdots$ defined on $(\Omega, \mathcal{F}, P)$ is *adapted* to the filtration if for each $n \geq 0$ the random variable M_n is measurable with respect to $\mathcal{F}_n$. An adapted sequence $(M_n)_{n\geq 0}$ is a *martingale* (respectively, *supermartingale*) if each M_n has finite first moment $E|M_n|$ and for any integers $m, n \geq 0$ and any event $F \in \mathcal{F}_m$,

$$E(M_{n+m}\mathbf{1}_F) = E(M_m\mathbf{1}_F) \quad \text{(martingale)} \tag{8.1.1}$$

$$E(M_{n+m}\mathbf{1}_F) \leq E(M_m\mathbf{1}_F) \quad \text{(supermartingale).} \tag{8.1.2}$$

Equivalently, the sequence $(M_n)_{n\geq 0}$ is a martingale (respectively, supermartingale) if for any integers $m, n \geq 0$,

$$M_m = E(M_{m+n} \,|\, \mathcal{F}_m) \quad \text{(martingale)} \tag{8.1.3}$$

$$M_m \geq E(M_{m+n} \,|\, \mathcal{F}_m) \quad \text{(supermartingale)} \tag{8.1.4}$$

See Section A.9 of the Appendix for the definition and basic properties of conditional expectation. Martingale theory is the invention of J. Doob [34], to whom not only the definition but also all of the main theorems of this chapter are due. Several of the examples we will consider (specifically, Examples 8.1.3 and 8.1.4) were studied earlier by A. Wald and P. Lévy.

If the sequence $(M_n)_{n\geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$, then for any integer $k \geq 0$ the sequence $(M_{n+k})_{n\geq k}$ is a martingale relative to the filtration $(\mathcal{F}_{n+k})_{n\geq 0}$, as is the sequence $(M_{n+k} - M_k)_{n\geq 0}$. Furthermore, since the σ-algebras in a filtration are nested, to verify the condition (8.1.1) it suffices to check the weaker condition

$$E(M_{m+1}\mathbf{1}_F) = E(M_m\mathbf{1}_F) \quad \forall\, F \in \mathcal{F}_m \text{ and } \forall\, m \geq 0. \tag{8.1.5}$$

The martingale identity (8.1.1), applied with $F = \Omega$, implies that for any martingale $(M_n)_{n\geq 0}$, "expectation is conserved": for every $n \geq 0$,

$$EM_n = EM_0. \tag{8.1.6}$$

Similarly, for any supermartingale $(M_n)_{n\geq 0}$,

$$EM_0 \geq EM_n \quad \text{for all } n \in \mathbb{N}. \tag{8.1.7}$$

Example 8.1.2 (Harmonic Martingales) Let $(X_n)_{n\geq 0}$ be a Markov chain on a finite or countable state space Θ, and let $h : \Theta \to \mathbb{R}_+$ be a nonnegative harmonic

function. For any $x \in \Theta$, if $E^x h(X_n) < \infty$ for all $n \in \mathbb{Z}_+$ then the sequence $M_n = h(X_n)$ is a martingale under any P^x relative to the filtration $(\mathcal{F}_n)_{n \geq 0}$, where

$$\mathcal{F}_n := \sigma(X_0, X_1, \cdots, X_n)$$

is the smallest σ-algebra containing all cylinder events

$$C(x_0, x_1, \cdots, x_n) = \bigcap_{i=0}^{n} \{X_i = x_i\}$$

of length $n + 1$. (Henceforth, we will call this the *standard filtration* for the Markov chain.) To prove this, it suffices to verify the martingale property (8.1.1) for $n = 1$ and cylinder events $F = C(x_0, x_1, \cdots, x_m)$. Since h is harmonic, we have, by the Markov property,

$$\begin{aligned}
E^x(h(X_{m+1})\mathbf{1}_F) &= \sum_{x_{m+1} \in \Theta} h(x_{m+1}) P^x(C(x_0, x_1, \cdots x_m, x_{m+1})) \\
&= P^x(C(x_0, x_1, \cdots, x_m)) \sum_{x_{m+1} \in \Theta} p(x_m, x_{m+1}) h(x_{m+1}) \\
&= P^x(C(x_0, x_1, \cdots, x_m)) h(x_m) \\
&= E^x(h(X_m)\mathbf{1}_F).
\end{aligned}$$

□

Example 8.1.3 If $(Y_n)_{n \geq 1}$ is a sequence of independent, identically distributed, real-valued random variables with finite first moment $E|Y_1|$ and mean $EY_1 = \alpha$, then the sequence

$$S_n := \sum_{i=1}^{n} (Y_i - \alpha) \tag{8.1.8}$$

is a martingale relative to the filtration $\mathcal{F}_n := \mathcal{F}(Y_1, Y_2, \cdots, Y_n)$. To see this, suppose without loss of generality that $\alpha = 0$, and let Z be a bounded random variable that is $\mathcal{F}_n$-measurable. Then by independence, $EY_{n+1}Z = EY_{n+1}EZ$, and so

$$\begin{aligned}
ES_{n+1}Z &= ES_nZ + EY_{n+1}Z \\
&= ES_nZ + EY_{n+1}EZ \\
&= ES_nZ.
\end{aligned}$$

A similar calculation shows that if the random variables Y_i also have finite *second* moment $EY_i^2 = \sigma^2 < \infty$ and mean $EY_i = 0$ then the sequence

$$S_n^2 - n\sigma^2 = \left(\sum_{i=1}^{n} Y_i\right)^2 - n\sigma^2 \tag{8.1.9}$$

is a martingale relative to the filtration $\mathcal{F}_n$.

Example 8.1.4 (Likelihood Ratio Martingales) Let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$ be a filtration of a measurable space $(\Omega, \mathcal{F})$, and let P, Q be probability measures on $\mathcal{F}$ such that for each $n \geq 0$ the measure Q is absolutely continuous with respect to P on the σ-algebra $\mathcal{F}_n$. Then the sequence of likelihood ratios (i.e., Radon-Nikodym derivatives; see Section A.8)

$$L_n := \left(\frac{dQ}{dP}\right)_{\mathcal{F}_n}$$

is a martingale (under P). The proof is trivial: any event $F \in \mathcal{F}_m$ is also an element of $\mathcal{F}_{m+1}$, so

$$\begin{aligned} Q(F) &= E_P(L_m \mathbf{1}_F) \quad (\text{since } F \in \mathcal{F}_m), \quad \text{and} \\ Q(F) &= E_P(L_{m+1} \mathbf{1}_F) \quad (\text{since } F \in \mathcal{F}_{m+1}). \end{aligned}$$

□

8.2 Doob's Optional Stopping Formula

The bedrock of martingale theory is the fact that the martingale and supermartingale properties are preserved under *optional stopping*. This fact leads to a far-reaching generalization of the Poisson formula (6.2.6) due to Doob known as the *Optional Stopping Formula*.

Definition 8.2.1 A random variable T valued in $\mathbb{Z}_+ \cup \{\infty\}$ is a *stopping time* for the filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ if for every nonnegative integer n the event $\{T = n\}$ is an element of the σ-algebra $\mathcal{F}_n$.

It is trivial that any constant random variable $T = m$ is a stopping time, and that the minimum of two stopping times is a stopping time. If $(X_n)_{n\geq 0}$ is a sequence of random variables that is *adapted* to the filtration, in the sense that for each n the random variable X_n is measurable with respect to $\mathcal{F}_n$, then for any sequence G_n of Borel sets the first entry time

$$T = \min\{n : X_n \in G_n\} \quad \text{or} \quad T = \infty \quad \text{it there is no such } n$$

is a stopping time. Similarly, if $(X_n)_{n\geq 0}$ is a Markov chain with state space Θ, then for any subset $U \subset \Theta$ the first exit time $\tau_U = \min\{n : X_n \notin U\}$ is a stopping time for the standard filtration.

Exercise 8.2.2 Show that if T_1 and T_2 are stopping times for a filtration $(\mathcal{F}_n)_{n\geq 0}$ then so are $T_1 \wedge T_2$ and $T_1 \vee T_2$.

Proposition 8.2.3 *If $(M_n)_{n\geq 0}$ is a martingale (respectively, supermartingale) and T a stopping time relative to a filtration $(\mathcal{F}_n)_{n\geq 0}$, then the sequence $(M_{n\wedge T})_{n\geq 0}$ is also a martingale (respectively, supermartingale) relative to $(\mathcal{F}_n)_{n\geq 0}$.*

Proof. The key is that for each m the event $\{T > m\}$ is measurable with respect to the σ-algebra $\mathcal{F}_m$, as it is the complement of $\cup_{n\leq m}\{T = n\}$. Thus, if the sequence $(M_n)_{n\geq 0}$ is a martingale then for any event $F \in \mathcal{F}_m$,

$$EM_{(m+1)}\mathbf{1}_F\mathbf{1}_{\{T>m\}} = EM_m\mathbf{1}_F\mathbf{1}_{\{T>m\}},$$

and so

$$\begin{aligned} EM_{(m+1)\wedge T}\mathbf{1}_F &= EM_{(m+1)\wedge T}\mathbf{1}_F\mathbf{1}_{\{T>m\}} + EEM_{(m+1)\wedge T}\mathbf{1}_F\mathbf{1}_{\{T\leq m\}} \\ &= EM_{(m+1)}\mathbf{1}_F\mathbf{1}_{\{T>m\}} + EEM_{m\wedge T}\mathbf{1}_F\mathbf{1}_{\{T\leq m\}} \\ &= EM_m\mathbf{1}_F\mathbf{1}_{\{T>m\}} + EEM_{m\wedge T}\mathbf{1}_F\mathbf{1}_{\{T\leq m\}} \\ &= EM_{m\wedge T}\mathbf{1}_F. \end{aligned}$$

This shows that the stopped sequence $(M_{n\wedge T})_{n\geq 0}$ is a martingale (cf. equation (8.1.5)). Essentially the same argument shows that if $(M_n)_{n\geq 0}$ is a supermartingale then so is $(M_{n\wedge T})_{n\geq 0}$. □

Corollary 8.2.4 (Doob's Optional Stopping Formula) *Let $(M_n)_{n\geq 0}$ be a martingale (respectively, supermartingale) and T a stopping time relative to a filtration $(\mathcal{F}_n)_{n\geq 0}$. Then for every nonnegative integer m*

$$EM_0 \overset{(\geq)}{=} EM_{T\wedge m}. \tag{8.2.1}$$

Proof. This follows directly from Proposition 8.2.3 and the "conservation of expectation" property (8.1.6) (and its relative (8.1.7) for supermartingales). □

The following example illustrates the utility of Doob's formula in random walk problems.

Proposition 8.2.5 *Let $(X_n)_{n\geq 0}$ be an irreducible random walk on a finitely generated group Γ with finite-index subgroup H, and let $(X_{T_m})_{m\geq 0}$ be the induced random walk on H (cf. Proposition 7.4.3). If the step distribution μ has finite first moment*

$$E|X_1| = \sum_{x \in \Gamma} |x|\mu(x) < \infty, \tag{8.2.2}$$

where $|x|$ denotes the word-norm for the group Γ, then the step distribution of the induced random walk on the subgroup H also has finite first moment, that is,

$$E^x|X_{T_1}^{-1} X_{T_2}| = E|X_T| < \infty. \tag{8.2.3}$$

Moreover, there is a constant $C < \infty$ depending only on the step distribution μ such that for every $x \in \Gamma$,

$$E^x|X_{T_1}| \leq |x| + C. \tag{8.2.4}$$

Proof. Let $T = T_1$ be the first return time to H. By Lemma 7.4.6, the distribution of $X_{T_1}^{-1} X_{T_2}$ under any P^x is the same as that of X_T under $P = P^1$, so to prove (8.2.3) it suffices to show that $E|X_T| < \infty$. By the subadditivity of the norm,

$$|X_T| \leq \sum_{i=1}^{T} |\xi_i|,$$

where $\xi_n := X_{n-1}^{-1} X_n$ are the steps of the random walk. Since $|\xi_1|$ has finite mean $E|\xi_1| := \alpha$, the sequence $S_n := \sum_{i=1}^{n} |\xi_i| - n\alpha$ is a martingale (cf. Example 8.1.3). Hence, by Doob's Formula, for every $m \in \mathbb{N}$,

$$E \sum_{i=1}^{T \wedge m} |\xi_i| = \alpha E(T \wedge m),$$

and so by the Monotone Convergence Theorem,

$$E \sum_{i=1}^{T} |\xi_i| = \alpha ET$$

By Lemma 7.4.5, $ET < \infty$, and so it follows that $E|X_T| \leq \alpha ET < \infty$. A similar argument shows that $E^z|X_T| < \infty$ for any initial point $z \in \Gamma$.

Finally, observe that any $x \in \Gamma$ can be written as $x = yx_i$ for some $y \in H$, where $(Hx_i)_{i \leq K}$ is an enumeration of the right cosets of H. Consequently, the distribution of X_T under P^x is the same as the distribution of yX_T under P^{x_i}, and so

$$E^x|X_T| = E^{x_i}|yX_T| \leq |y| + E^{x_i}|X_T|.$$

The inequality (8.2.4) follows, with $C = \max_{i \leq K} E^{x_i}|X_T|$. □

Exercise 8.2.6 Show that the hypothesis of finite support in the first assertion of Lemma 7.4.9 can be weakened to finite second moment.

HINT: Apply Doob's identity, but now for the martingale

$$V_n := \left(\sum_{i=1}^{n} |\xi_i| - nE|\xi_1|\right)^2 .$$

8.3 The Martingale Convergence Theorem

Theorem 8.3.1 (Martingale Convergence Theorem) *If $(M_n)_{n\geq 0}$ is a nonnegative supermartingale then*

$$M_\infty := \lim_{n\to\infty} M_n \tag{8.3.1}$$

exists and is finite with probability 1. *If $(M_n)_{n\geq 0}$ is a* uniformly bounded *martingale on the filtration $(\mathcal{F}_n)_{n\geq 0}$ (that is, if there exists $C \in \mathbb{R}$ such that $|M_n| \leq C$ for every $n \in \mathbb{Z}_+$), then for every $m \geq 0$ and every event $F \in \mathcal{F}_m$,*

$$EM_m\mathbf{1}_F = EM_\infty\mathbf{1}_F. \tag{8.3.2}$$

For *martingales*, almost sure convergence holds under the less restrictive condition that $\sup_{n\geq 0} E|M_n| < \infty$. Since we will only have occasion to use the result for *nonnegative* martingales, we will not prove this more general theorem (which can be found in Doob [34]).

The last assertion (8.3.2) for uniformly bounded martingales follows directly from almost sure convergence and the martingale property, by the bounded convergence theorem for integrals, because the martingale property ensures that $EM_{m+n}\mathbf{1}_F = EM_m\mathbf{1}_F$ for every $n \geq 0$. The theorem therefore implies that if $\mathcal{F}_\infty$ is the smallest σ-algebra containing every $\mathcal{F}_n$ then for every bounded, $\mathcal{F}_\infty$-measurable random variable M_∞,

$$M_\infty = \lim_{n\to\infty} M_n = \lim_{n\to\infty} E(M_\infty \,|\, \mathcal{F}_n) \quad \text{almost surely.} \tag{8.3.3}$$

For unbounded martingales, the convergence (8.3.2) of expectations need not hold. A simple example is the *double-or-nothing*, or *Saint Petersburg martingale*, defined by

$$M_{n+1} = 2M_n\xi_{n+1} \quad \text{and} \quad M_0 = 1 \tag{8.3.4}$$

where $\xi_1, \xi_2, \cdots$ are independent, identically distributed Bernoulli$-(1/2)$ random variables. Here $M_\infty = 0$ almost surely, so (8.3.2) fails for $F = \Omega$. Another example

is the harmonic martingale $h_\theta(X_n)$ where X_n is the simple random walk on the free group $\mathbb{F}_2$ and h_θ is the harmonic function in Example 6.2.3.

Example 8.1.2 shows that Theorem 8.0.1 is a special case of Theorem 8.3.1. Theorem 8.3.1 also has interesting consequences for the likelihood ratio martingale L_n described in Example 8.1.4. First, it implies that the sequence $(L_n)_{n\geq 0}$ of likelihood ratios converges almost surely under P. The limit random variable L_∞ must be measurable with respect to the smallest σ-algebra $\mathcal{F}_\infty$ containing all of the σ-algebras $\mathcal{F}_n$ of the filtration, and by Fatou's lemma,

$$E_P L_\infty \leq \lim_{n\to\infty} E_P L_n = 1.$$

If equality holds, then the dominated convergence theorem implies that for every event $F \in \cup_{n=0}^\infty \mathcal{F}_n$,

$$E_P(L_\infty \mathbf{1}_F) = \lim_{n\to\infty} E_P(L_n \mathbf{1}_F) = Q(F)$$

which in turn implies that Q is absolutely continuous with respect to Pon $\mathcal{F}_\infty$, with likelihood ratio L_∞. This must be the case, in particular, if the likelihood ratio sequence L_n is (uniformly) bounded.

The proof of almost sure convergence for nonnegative supermartingales will rely on two basic martingale inequalities, the *maximal* and the *upcrossings* inequalities.

Proposition 8.3.2 (Maximal Inequality) *If $(M_n)_{n\geq 0}$ is a nonnegative supermartingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$, then for any real numbers $\beta > \alpha > 0$, and any event $G \in \mathcal{F}_0$,*

$$P\left(\left\{\sup_{n\geq 0} M_n \geq \beta\right\} \cap \{M_0 \leq \alpha\} \cap G\right) \leq (\alpha/\beta) P(\{M_0 \leq \alpha\} \cap G). \tag{8.3.5}$$

Proof. It suffices (as you should check) to prove the inequality with the event $\left\{\sup_{n\geq 0} M_n \geq \beta\right\}$ replaced by $\left\{\sup_{n\geq 0} M_n > \beta\right\}$. Write $F = \{M_0 \leq \alpha\}$; this event is an element of $\mathcal{F}_0$, and hence of every $\mathcal{F}_n$. Define $T = T(\beta)$ to be the smallest $n \geq 0$ for which $M_n > \beta$, or $T = \infty$ if no such n exists. Clearly,

$$\left\{\sup_{n\geq 0} M_n > \beta\right\} = \{T(\beta) < \infty\}.$$

The random variable T is a stopping time, so Proposition 8.2.3 applies; thus, we have

$$\begin{aligned}
\beta P(\{T \leq m\} \cap F \cap G) &\leq E(M_{T\wedge m} \mathbf{1}_{\{T\leq m\}} \mathbf{1}_{F\cap G}) \\
&\leq E(M_{T\wedge m} \mathbf{1}_{F\cap G}) \\
&\leq E(M_0 \mathbf{1}_{F\cap G})
\end{aligned}$$

$$\leq \alpha P(F \cap G),$$

Hence, by monotone convergence,

$$\beta P(\{T < \infty\} \cap F \cap G) = \beta P\left(\left\{\sup_{n\geq 0} M_n > \beta\right\} \cap F \cap G\right) \leq \alpha P(F \cap G).$$

□

Corollary 8.3.3 *For any nonnegative supermartingale* $(M_n)_{n\geq 0}$ *and any real number* $\beta > 0$*,*

$$P\left\{\sup_{n\geq 0} M_n \geq \beta\right\} \leq \frac{EM_0}{\beta}. \tag{8.3.6}$$

Proof. Without loss of generality, we may assume that the supermartingale has constant initial term $M_0 = EM_0$, because if not, the extended sequence $EM_0, M_0, M_1, \cdots$ is also a supermartingale (with respect to the filtration obtained by prepending the trivial σ-algebra $\mathcal{F}_0 = \{\Omega, \emptyset\}$ to the original filtration). Applying the maximal inequality (8.3.5) with $\alpha = EM_0 = M_0$ and $G = \Omega$ yields (8.3.6). □

Corollary 8.3.3 implies that any nonnegative supermartingale $(M_n)_{n\geq 0}$ has finite supremum, with probability 1. Therefore, except on an event of probability 0, the sequence $(M_n)_{n\geq 0}$ will fail to converge to a finite limit only if there is a nonempty interval $(\alpha, \beta) \subset (0, \infty)$ with rational endpoints such that

$$\liminf_{n\to\infty} M_n \leq \alpha \quad \text{and}$$

$$\limsup_{n\to\infty} M_n \geq \beta.$$

This will occur only if the sequence makes infinitely many *upcrossings* from $[0, \alpha]$ to $[\beta, \infty)$. The following inequality shows that this cannot happen with positive probability.

Proposition 8.3.4 (Upcrossings Inequality) *Let* $(M_n)_{n\geq 0}$ *be a nonnegative supermartingale. Fix* $0 \leq \alpha < \beta < \infty$*, and let* N *be the number of* upcrossings *of* $[\alpha, \beta]$*, that is, the number of times that the sequence* M_n *crosses from below* α *to above* β*. Then for every* $m = 0, 1, 2, \cdots$*,*

$$P\{N \geq m\} \leq \left(\frac{\alpha}{\beta}\right)^m. \tag{8.3.7}$$

Proof. This is by induction on m. The inequality is trivial if $m = 0$, so it suffices to show that if it is true for m then it is true for $m + 1$.

The event $N \geq m+1$ occurs only if $N \geq m$ and the sequence M_n returns to the interval $[0, \alpha]$ following the completion of its mth upcrossing. Define τ_m to be the time n of the first return to $[0, \alpha]$ after the mth upcrossing (or $\tau_m = \infty$ if either the mth upcrossing is never completed or if there is no return to $[0, \alpha]$ afterwards), and define T_{m+1} to be the first $n > \tau_m$ such that $M_n \geq \beta$. Obviously, the event $\{N \geq m+1\}$ coincides with $\{T_{m+1} < \infty\}$.

For any integer $k \geq 0$ the sequence $(M_{n+k})_{n \geq 0}$ is a supermartingale with respect to the filtration $(\mathcal{F}_{n+k})_{n \geq 0}$. The event $G = \{\tau_m = k\}$ is an element of the σ-algebra $\mathcal{F}_k$, and G is contained in the event $\{M_k \leq \alpha\}$, so by the Maximal Inequality,

$$P\left(\left\{\sup_{n \geq k} M_n \geq \beta\right\} \cap \{\tau_m = k\}\right) \leq (\alpha/\beta) P\{\tau_m = k\}.$$

Thus,

$$P\left(\left\{\sup_{n \geq \tau_m} M_n \geq \beta\right\} \cap \{\tau_m < \infty\}\right) \leq (\alpha/\beta) P\{\tau_m < \infty\} \leq (\alpha/\beta) P\{N \geq m\}.$$

The event $\{N \geq m+1\}$ coincides with $\left\{\sup_{n \geq \tau_m} M_n \geq \beta\right\} \cap \{\tau_m < \infty\}$, so this proves that

$$P\{N \geq m+1\} \leq (\alpha/\beta) P\{N \geq m\},$$

completing the induction. □

Proof of Theorem 8.3.1 The Maximal Inequality implies that $\sup_{n \geq 0} M_n$ is finite with probability 1, and the Upcrossings Inequality implies that for any two real numbers $\alpha < \beta$, the number of crossings of the interval (α, β) is almost surely finite. Because the set of *rational* pairs $\alpha < \beta$ is countable, and because the union of countably many sets of probability 0 is a set of probability 0, it follows that with probability 1 there is no rational pair $\alpha < \beta$ such that the sequence M_n makes infinitely many upcrossings of the interval (α, β). Thus, with probability 1,

$$\liminf_{n \to \infty} M_n = \limsup_{n \to \infty} M_n$$

and so the sequence M_n must converge to a finite limit. □

Exercise 8.3.5 (*Martingales with bounded increments.*) Let $\{S_n\}_{n \geq 0}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n \geq 0}$ such that $S_0 = 0$ and such that the *martingale differences* $\xi_n := S_n - S_{n-1}$ are uniformly bounded. Assume for convenience that $|\xi_n| \leq 1$.

(A) For any integer $m \geq 1$ define

$$\tau_m = \min\{n \geq 0 : S_n \leq -m\},$$

$$= +\infty \quad \text{if } S_{-n} > -m \; \forall n \in \mathbb{N}.$$

Prove that $\lim_{n\to\infty} S_{n\wedge\tau_m} := Z_m$ exists and is finite almost surely.

HINT: Every *nonnegative* martingale converges almost surely.

(B) Conclude from (A) that on the event $\{\inf_{n\geq 1} S_n > -\infty\}$, the limit $\lim_{n\to\infty} S_n$ exists and is finite almost surely. Deduce that on the event $\{\sup_{n\geq 1} S_n < \infty\}$, the limit $\lim_{n\to\infty} S_n$ exists and is finite almost surely.)

(C) Now conclude that with probability one, *either* $\lim_{n\to\infty} S_n$ exists and is finite *or* $\limsup_{n\to\infty} S_n = -\liminf_{n\to\infty} S_n = \infty$.

(D) Let $\{A_n\}_{n\geq 1}$ be a sequence of events such that $A_n \in \mathcal{F}_n$ for every $n \geq 1$. Prove that the events

$$\left\{\sum_{n=1}^{\infty} \mathbf{1}_{A_n} = \infty\right\} \quad \text{and} \quad \left\{\sum_{n=1}^{\infty} P(A_n \,|\, \mathcal{F}_{n-1}) = \infty\right\}$$

differ by at most an event of probability 0.

HINT: Deduce this from (C). You will need to find an appropriate martingale with bounded increments.

NOTE: This generalizes the usual Borel-Cantelli Lemma. It is quite a useful generalization because it does not require the events A_n to be independent.

8.4 Martingales and Harmonic Functions†

In classical potential theory, a *harmonic function* is a real-valued function $u : \mathcal{W} \to \mathbb{R}$ defined in an open subset $\mathcal{W}$ of $\mathbb{R}^d$ that satisfies the *mean value property*: for every point $z \in U$ and every ball $\mathbb{B}_r(z) \subset \mathcal{W}$ centered at z that is contained in $\mathcal{W}$, the value $u(z)$ is the average of the values $u(z')$ over the ball $z' \in \mathbb{B}_r(z)$, that is,

$$u(z) = \frac{1}{|\mathbb{B}_r(z)|} \int_{\mathbb{B}_r(z)} u(z')\, dz'. \tag{8.4.1}$$

In the 1940s S. Kakutani [71] (following earlier work by Courant, Friedrichs, and Lewy [29]) discovered a fundamental connection between the boundary behavior of harmonic functions and the theory of d-dimensional *Brownian motion*. In modern language, if u is a (sufficiently nice) harmonic function and $(W_t)_{t\geq 0}$ is a Brownian motion (which we will not define here), then the composition $(u(W_t))_{t\geq 0}$ is a continuous-time martingale. Continuous-time random processes are beyond the scope of these notes; however, some of Kakutani's theory applies to a related discrete-time stochastic process, which we will call a *Brownian random walk*.

Definition 8.4.1 Let $\mathcal{W}$ be a bounded, connected, open set in $\mathbb{R}^d$, and let $(\Omega, \mathcal{F}, P)$ be a probability space on which are defined independent, identically distributed

random variables $Z_1, Z_2, \cdots$ uniformly distributed on the unit ball $\mathbb{B}_1$ in $\mathbb{R}^d$, i.e.,

$$P\{Z_i \in F\} = \frac{|\mathbb{B}_1 \cap F|}{\pi}; \tag{8.4.2}$$

here $|\cdot|$ denotes two-dimensional Lebesgue measure. A *Brownian random walk* in $\mathcal{W}$ with jump parameter $\varepsilon \in (0, 1)$ and initial point $W_0^z = z \in \mathcal{W}$ is the sequence $(W_n^z)_{n\geq 0}$ defined inductively by

$$W_{n+1}^z = W_n^z + \varepsilon\varrho(W_n^z)Z_{n+1} \tag{8.4.3}$$

where $\varrho(w)$ is the *inradius* of the region $\mathcal{W}$ at w, that is, the maximal real number $r > 0$ such that the open ball of radius r centered at w is contained in the closure $\mathcal{W} \cup \partial\mathcal{W}$.

Exercise 8.4.2 Assume that $\mathcal{W}$ is a bounded, connected open subset of $\mathbb{R}^d$, and let $(W_n^z)_{n\geq 0}$ be the Brownian random walk in $\mathcal{W}$ with initial point z.

(A) Check that if $u : \mathcal{W} \to \mathbb{R}_+$ is a nonnegative harmonic function in $\mathcal{W}$ then for any $z \in \mathcal{W}$ the sequence $u(W_n^z)$ is a martingale relative to the standard filtration $\mathcal{F}_n = \sigma(Z_1, Z_2, \cdots, Z_n)$.

(B) Check that each of the coordinate functions on $\mathcal{W}$ is harmonic. Also check that for any two indices $i, j \in [d]$ the function $u(x) = x_i^2 - x_j^2$ is harmonic.

(C) Prove that if the domain $\mathcal{W}$ is bounded, then for each $z \in \mathcal{W}$

$$\lim_{n\to\infty} W_n^z := W_\infty^z \in \partial\mathcal{W} \tag{8.4.4}$$

exists with probability one. Denote by ν_z the distribution of the limit point W_∞^z; call this the *exit measure*.

(D) Show that if $u : \mathcal{W} \to \mathbb{R}$ is a harmonic function that has a continous extension to the closure $\bar{\mathcal{W}}$ of $\mathcal{W}$, then for every $z \in \mathcal{W}$,

$$u(z) = \int_{\partial\mathcal{W}} u(x)\, d\nu_z(x). \tag{8.4.5}$$

This is the *Poisson integral formula* of classical potential theory.

(E)** Let $d = 2$, and assume that the boundary $\partial\mathcal{W}$ is a smooth (at least twice continuous differentiable) closed curve. Show that for any sequence of points $z_n \in \mathcal{W}$ that converges to a point $z \in \partial\mathcal{W}$ the exit measures ν_{z_n} converge *weakly* to ν_z.

HINTS: This is equivalent to showing that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \nu_{z_n}(\mathbb{B}_\varepsilon(z)) = 1.$$

For this, use the maximal inequality for the martingales $u(W_n^z)_{n\geq 0}$, where u is any one of the harmonic functions identified in part (B).

8.5 Reverse Martingales

Definition 8.5.1 A *reverse filtration* of a probability space $(\Omega, \mathcal{F}, P)$ is a *decreasing* sequence of σ-algebras $\mathcal{F} = \mathcal{F}_0 \supset \mathcal{F}_1 \supset \cdots$ all contained in $\mathcal{F}$. A *reverse martingale* with respect to the reverse filtration $(\mathcal{F}_n)_{n\geq 0}$ is a sequence of integrable random variables such that

(a) each random variable M_n is measurable with respect to $\mathcal{F}_n$; and
(b) for any $m, n \geq 0$ and any event $F \in \mathcal{F}_{m+n}$,

$$E(M_{n+m}\mathbf{1}_F) = E(M_m\mathbf{1}_F). \tag{8.5.1}$$

Example 8.5.2 Let $(X_n)_{n\geq 0}$ be a nearest-neighbor Markov chain with transition probability matrix p on a state space Θ, and let $w : \Theta \times \mathbb{Z}_+ \to \mathbb{R}_+$ be a nonnegative function such that for every $x \in \Theta$ and $n \in \mathbb{Z}_+$,

$$w(x, n+1) = \sum_{y\in\Theta} p(x, y)w(y, n). \tag{8.5.2}$$

Call such a function *reverse space-time harmonic*. If w is reverse space-time harmonic then the sequence $M_n = w(X_n, n)$ is a reverse martingale, under any P^x, relative to the reverse filtration $(\mathcal{F}_n)_{n\geq 0}$, where

$$\mathcal{F}_n = \sigma(X_n, X_{n+1}, X_{n+2}, \cdots)$$

is the smallest σ-algebra with respect to which all of the random variables $\{X_k\}_{k\geq n}$ are measurable. The proof is similar to that of Example 8.1.2.

Example 8.5.3 Let $(\mathcal{F}_n)_{n\geq 0}$ be a reverse filtration, and let P, Q be two probability measures on $\mathcal{F} = \mathcal{F}_0$ such that Q is absolutely continuous with respect to P on $\mathcal{F}$. This implies that Q is absolutely continuous with respect to P on every $\mathcal{F}_n$, so the Radon-Nikodym theorem implies that for each $n \geq 0$ there is a likelihood ratio

$$L_n := \left(\frac{dQ}{dP}\right)_{\mathcal{F}_n}. \tag{8.5.3}$$

The sequence L_n is a reverse martingale with respect to the reverse filtration $(\mathcal{F}_n)_{n\geq 0}$.

Example 8.5.4 † Let $Y_1, Y_2, \cdots$ be a sequence of independent, identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. If $E|Y_1| < \infty$

then the sequence

$$A_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$$

is a reverse martingale relative to the reverse filtration $(\mathcal{G}_n)_{n\geq 1}$, where

$$\mathcal{G}_n = \sigma(A_n, A_{n+1}, A_{n+2}, \cdots).$$

Exercise 8.5.5 † Prove this.

HINT: It suffices to show that for any Borel set $B \subset \mathbb{R}$ and any $i \in [n]$

$$EY_i\mathbf{1}_B(A_n) = EY_1\mathbf{1}_B(A_n).$$

Theorem 8.5.6 *If* $(M_n)_{n\geq 0}$ *is a nonnegative reverse martingale with respect to a reverse filtration* $(\mathcal{F}_n)_{n\geq 0}$ *then*

$$\lim_{n\to\infty} M_n := M_\infty \tag{8.5.4}$$

exists and is finite almost surely. The limit random variable M_∞ *is measurable with respect to the* σ*-algebra* $\mathcal{F}_\infty := \cap_{n\geq 0}\mathcal{F}_n$*. Furthermore, the convergence* $M_n \to M_\infty$ *also holds in the* L^1*-metric, that is,*

$$\lim_{n\to\infty} E|M_n - M_\infty| = 0. \tag{8.5.5}$$

Note: In fact, the hypothesis of nonnegativity is superfluous, but the only reverse martingales that we will encounter in these lectures (in Sections 9.7 and 10.4 below) will be nonnegative, so we will only prove the result for these.

Proof of Almost Sure Convergence. If $(M_n)_{n\geq 0}$ is a reverse martingale relative to the reverse filtration $(\mathcal{F}_n)_{n\geq 0}$, then for any integer $k \geq 0$ the sequence

$$M_k, M_{k-1}, \cdots, M_1, M_0, M_0, M_0, \cdots$$

is a *martingale* with respect to the filtration

$$\mathcal{F}_k, \mathcal{F}_{k-1}, \cdots, \mathcal{F}_1, \mathcal{F}_0, \mathcal{F}_0, \mathcal{F}_0, \cdots.$$

Consequently, for any rational numbers $0 \leq \alpha < \beta$ the number N_k of upcrossings of the interval $[\alpha, \beta]$ by this sequence satisfies

$$P\{N_k \geq m\} \leq \left(\frac{\alpha}{\beta}\right)^m. \tag{8.5.6}$$

As $k \to \infty$, the random variables N_k increase to $N =$ the number of *downcrossings* of $[\alpha, \beta]$ by the infinite sequence $(M_n)_{n \geq 0}$. Thus, with probability one, the number of downcrossings of any rational interval $[\alpha, \beta]$ is finite. It follows that with probability one the sequence M_n converges either to a finite limit or to $+\infty$. Since the expectations EM_n are all equal and finite, the event $\{\lim_{n\to\infty} M_n = \infty\}$ must have probability 0, so the limit (8.5.4) holds almost surely, and the limit is finite. The limit random variable M_∞ is measurable with respect to any $\mathcal{F}_n$ in the reverse filtration, since $M_\infty = \limsup_{m\to\infty} M_{n+m}$, so M_∞ must be measurable with respect to $\mathcal{F}_\infty$. □

Proof of L^1 Convergence We can assume without loss of generality that $EM_0 = 1$. (If not, replace the original reverse martingale M_n by $M_n' = M_n/EM_0$.) Let Q be the probability measure on $\mathcal{F} = \mathcal{F}_0$ with likelihood ratio M_0:

$$Q(F) = E(M_0 \mathbf{1}_F) \quad \forall\, F \in \mathcal{F}.$$

(Here E denotes expectation with respect to P.) Because Radon-Nikodym derivatives are unique (up to change on events of probability 0; cf. Theorem A.8.6), the terms M_n of the reverse martingale coincide with the likelihood ratios $L_n = (dQ/dP)_{\mathcal{F}_n}$, that is, for every n,

$$M_n = \left(\frac{dQ}{dP}\right)_{\mathcal{F}_n} = L_n \quad \text{almost surely.}$$

Furthermore (see Exercise A.8.5 of the Appendix), since Q is absolutely continuous with respect to P on $\mathcal{F}$, for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any event $F \in \mathcal{F}$,

$$P(F) < \delta \quad \Longrightarrow \quad Q(F) < \varepsilon. \tag{8.5.7}$$

Set $L_\infty = (dQ/dP)_{\mathcal{F}_\infty}$ and $M_\infty = \lim_{n\to\infty} M_n = \lim_{n\to\infty} L_n$; observe that both L_∞ and M_∞ are measurable relative to $\mathcal{F}_\infty$. For any event $F \in \mathcal{F}_\infty$,

$$\lim_{n\to\infty} L_n \mathbf{1}_F = M_\infty \mathbf{1}_F \quad \text{almost surely} \quad \Longrightarrow$$
$$EM_\infty \mathbf{1}_F \leq \liminf_{n\to\infty} EL_n \mathbf{1}_F = Q(F) = EL_\infty \mathbf{1}_F,$$

by Fatou's Lemma. Consequently, $M_\infty \leq L_\infty$ almost surely.

Fix $\varepsilon > 0$, and for each $n \in \mathbb{N}$ define

$$F_n = \{M_\infty - M_n > \varepsilon\} \quad \text{and}$$
$$G_n = \{M_\infty \leq M_n - \varepsilon\}.$$

Since $\lim M_n = M_\infty$ almost surely, $\lim P(F_n) = \lim P(G_n) = 0$, and so by (8.5.7), $\lim Q(F_n) = \lim Q(G_n) = 0$. Thus, since $M_\infty \le L_\infty$ almost surely,

$$\begin{aligned}\limsup_{n\to\infty} E(M_\infty - M_n)_+ &\le \varepsilon + \limsup_{n\to\infty} E(M_\infty - M_n)_+ \mathbf{1}_{F_n} \\ &\le \varepsilon + \limsup_{n\to\infty} EM_\infty \mathbf{1}_{F_n} \\ &\le \varepsilon + \limsup_{n\to\infty} EL_\infty \mathbf{1}_{F_n} = \varepsilon,\end{aligned}$$

the last by the dominated convergence theorem (as L_∞ is integrable). Similarly,

$$\begin{aligned}\limsup_{n\to\infty} E(M_\infty - M_n)_- &\le \varepsilon + \limsup E(M_\infty - M_n)_- \mathbf{1}_{G_n} \\ &\le \varepsilon + \limsup EM_n \mathbf{1}_{G_n} \\ &\le \varepsilon + \limsup_{n\to\infty} Q(G_n) = \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ can be chosen arbitrarily small, it follows that $\lim_{n\to\infty} E|M_n - M_\infty| = 0$, proving the L^1-convergence, and incidentally showing that $M_\infty = L_\infty$ almost surely. □

Corollary 8.5.7 *Let $(\mathcal{F}_n)_{n\ge 0}$ be a reverse filtration and let P, Q be probability measures on $\mathcal{F} = \mathcal{F}_0$ such that Q is absolutely continuous with respect to P on $\mathcal{F}$. Let L_n be the sequence of likelihood ratios* (8.5.3). *Then*

$$\lim_{n\to\infty} L_n := L_\infty = \left(\frac{dQ}{dP}\right)_{\mathcal{F}_\infty} \quad \textit{almost surely and in } L^1, \tag{8.5.8}$$

where $\mathcal{F}_\infty := \cap_{n\ge 0}\mathcal{F}_n$.

Proof. The sequence L_n is a reverse martingale (Example 8.5.3), so the reverse martingale convergence theorem implies that it has an almost sure limit L_∞. □

Additional Notes. Martingales are introduced here primarily so that we will have access to the basic convergence theorems (Theorems 8.3.1 and 8.5.6) in later chapters. However, martingales enter into random walk theory in a variety of other ways. The *Martingale Central Limit Theorem* is an essential tool in establishing approximate normality of various functionals of interest: see [83] for a textbook introduction, or [64] for an extensive treatment. *Azuma's Inequality* is a generalization of Hoeffding's Inequality (Proposition A.6.1) to martingale difference sequences; it is fundamental to the study of *concentration inequalities*. Once again, see [83] for an introduction.

The brief discussion in Section 8.4 only hints at the importance of martingales in the study of harmonic functions (as classically defined). See Bass [7] and Mörters and Peres [100] for textbook introductions to the deep connections between martingale theory, Brownian motion, and harmonic functions.

Chapter 9
Bounded Harmonic Functions

9.1 The Invariant σ-Algebra $\mathcal{I}$

The Martingale Convergence Theorem (cf. Theorems 8.0.1 and 8.3.1) implies that for any Markov chain $\mathbf{X} = (X_n)_{n\geq 0}$ on a countable set Θ and any nonnegative harmonic function $u : \Theta \to \mathbb{R}$ the sequence $u(X_n)$ converges P^x-almost surely to a finite limit Y, for any initial point $x \in \Theta$. Consequently, by the dominated convergence theorem, if a harmonic function u is *bounded*, then it has the integral representation

$$u(x) = E^x Y. \tag{9.1.1}$$

Which random variables Y can arise as limits? Is the integral representation (9.1.1) unique? Our objective in the first two sections of this chapter will be to explore aspects of the limit theory of Markov chains that bear on these questions.

If $Y = \lim_{n\to\infty} u(X_n)$ then the limit random variable Y is a function of the sequence $(X_n)_{n\geq 0}$, but it doesn't depend on the entire sequence: in fact, for any $m \geq 1$ the value of Y is determined by the truncated sequence $X_{m+1}X_{m+2}, \cdots$. Since m is arbitrary, it follows that Y is a *shift-invariant* function of the sequence X_n. Let's begin by understanding what this means.

Recall that a Markov chain $\mathbf{X} = (X_n)_{n\geq 0}$ with state space Θ can be viewed as a random variable $\mathbf{X} : \Omega \to \Theta^\infty$ taking values in the sequence space Θ^∞. (Here $(\Omega, \mathcal{F}, P^x)$ is the probability space[1] on which the Markov chain is defined.) The *law* (or *joint distribution*) of a Markov chain $\mathbf{X}$ with initial point x (cf. Definition 6.1.7) is the induced measure $\hat{P}^x$ on the Borel σ-algebra $\mathcal{B}_\infty$ of Θ^∞ defined by

[1] It is a minor nuisance to have to distinguish between the probability spaces $(\Omega, \mathcal{F}, P^x)$ and $(\Theta^\infty, \mathcal{B}_\infty, \hat{P}^x)$. There are, however, good reasons for allowing Markov chains to be defined on arbitrary probability spaces, as this allows for constructions that require auxiliary random variables. In particular, it allows *coupling* constructions such as those in Section 9.6 below.

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_9

$$\hat{P}^x := P^x \circ \mathbf{X}^{-1}; \tag{9.1.2}$$

if $\hat{X}_0, \hat{X}_1, \dots : \Theta^\infty \to \Theta$ are the coordinate projections, then under $\hat{P}^x$ the sequence $(\hat{X}_n)_{n\geq 0}$ is a version of the Markov chain $(X_n)_{n\geq 0}$. Denote by $\sigma : \Theta^\infty \to \Theta^\infty$ the (forward) shift mapping on Θ^∞:

$$\sigma(x_0, x_1, x_2, \cdots) = (x_1, x_2, x_3, \cdots). \tag{9.1.3}$$

Definition 9.1.1 The *invariant σ-algebra $\mathcal{I}$* on Θ^∞ is the set of all events $F \in \mathcal{B}_\infty$ whose indicators satisfy $\mathbf{1}_F = \mathbf{1}_F \circ \sigma$, or equivalently,

$$(x_0, x_1, x_2, \cdots) \in F \iff (x_1, x_2, x_3, \cdots) \in F. \tag{9.1.4}$$

Exercise 9.1.2 Let $T : \Theta \to \mathcal{Y}$ be a function taking values in a metric space $\mathcal{Y}$, and let $z_* \in \mathcal{Y}$ be a fixed element of $\mathcal{Y}$. Show that the function $Z : \Theta^\infty \to \mathcal{Y}$ defined by

$$Z((x_0, x_1, \cdots)) = \lim_{n\to\infty} T(x_n) \quad \text{if the limit exists, and}$$
$$Z((x_0, x_1, \cdots)) = z_* \quad \text{otherwise}$$

is $\mathcal{I}$-measurable (relative to the Borel σ-algebra $\mathcal{B}_\mathcal{Y}$ on $\mathcal{Y}$). Deduce from this that for any nonnegative, superharmonic function $u : \Theta \to \mathbb{R}_+$ for the Markov chain $(X_n)_{n\geq 0}$, there exists an $\mathcal{I}$-measurable function $\psi : \Theta^\infty \to \mathbb{R}_+$ such that for every $x \in \Theta$,

$$\psi(X_0, X_1, \cdots) = \lim_{n\to\infty} u(X_n) \quad P^x - \text{almost surely.}$$

NOTE: The *Borel σ-algebra $\mathcal{B}_\mathcal{Y}$* on a topological space $\mathcal{Y}$ is defined to be the smallest σ-algebra containing all the open sets of $\mathcal{Y}$. If $\mathcal{Y}$ is a metric space then $\mathcal{B}_\mathcal{Y}$ is the smallest σ-algebra containing all the open balls. See Section A.7 of the Appendix.

Definition 9.1.3 For any initial state $x \in \Theta$, define the *exit measure* of the Markov chain for the initial state x to be the restriction $\nu_x = \hat{P}^x \restriction \mathcal{I}$ of the measure $\hat{P}^x$ to the invariant σ-algebra $\mathcal{I}$.

The specification of the σ-algebra is critically important: the following proposition does not hold for the measures $\hat{P}^x$ on the full Borel σ-algebra $\mathcal{B}_\infty$.

Proposition 9.1.4 *The exit measures satisfy the* mean value property

$$\nu_x = \sum_{y\in\Theta} p(x, y)\nu_y \quad \textit{for every } x \in \Theta. \tag{9.1.5}$$

Consequently, for every bounded, $\mathcal{I}$-measurable random variable $Y : \Theta^\infty \to \mathbb{R}$ the function $u : \Theta \to \mathbb{R}$ defined by

$$u(x) = E^x(Y \circ \mathbf{X}) = \hat{E}^x Y = \int Y \, d\nu_x \tag{9.1.6}$$

is bounded and harmonic.

Proof. The Markov property implies that for any invariant event F and any point $x \in \Theta$,

$$\begin{aligned}
\nu_x(F) &= P^x \{(X_0, X_1, X_2, \cdots) \in F\} \\
&= P^x \{(X_1, X_2, X_3, \cdots) \in F\} \\
&= \sum_{y \in \Theta} p(x, y) P^y \{(X_0, X_1, \cdots) \in F\} \\
&= \sum_{y \in \Theta} p(x, y) \nu_y(F).
\end{aligned}$$

This proves (9.1.5), and also shows that the function $x \mapsto \nu_x(F)$ is harmonic. Since harmonicity is preserved by linear combinations and limits, the second assertion follows. □

Equation (9.1.6) defines a linear mapping from the space of bounded, $\mathcal{I}$-measurable random variables to the space of bounded harmonic functions for the Markov chain. The following theorem of D. Blackwell [13] asserts that this mapping is a bijection.

Theorem 9.1.5 (Blackwell) *Every bounded harmonic function u has a representation* (9.1.6) *for some bounded, $\mathcal{I}$-measurable random variable Y, and for every $x \in \Theta$,*

$$Y \circ \mathbf{X} = \lim_{n \to \infty} u(X_n) \quad P^x - \textit{almost surely.} \tag{9.1.7}$$

Consequently, the representation (9.1.6) *is essentially unique, in the following sense. If Y_1, Y_2 are two bounded, $\mathcal{I}$-measurable random variables on Θ^∞ such that $\hat{E}^x Y_1 = \hat{E}^x Y_2$ for every state $x \in \Theta$ then $\nu_x \{Y_1 \neq Y_2\} = 0$ for every x.*

Proof. Proposition 9.1.4 implies that if Y is a bounded, invariant (that is, $\mathcal{I}$-measurable) random variable then the function $u(x) = \hat{E}^x Y$ is harmonic. On the other hand, the Martingale Convergence Theorem implies that for any bounded harmonic function u the limit $Z = \lim_{n \to \infty} u(\hat{X}_n)$ exists $\hat{P}^x$-almost surely, and so the Dominated Convergence Theorem implies that u has the integral representation $u(x) = \hat{E}^x Z$. By Exercise 9.1.2, Z must be $\mathcal{I}$-measurable. This proves that every bounded harmonic function is of the form (9.1.6).

To prove (9.1.7), we must show that if Y is a bounded, invariant random variable and $u(x) := E^x Y$, then the almost sure limit $Z := \lim_{n\to\infty} u(\hat{X}_n)$, which exists by the Martingale Convergence Theorem, equals Y almost surely. By the Dominated Convergence Theorem, $E^x Z = E^x Y$ for every $x \in \Theta$; consequently, since $Z - Y$ is a bounded and invariant, it will suffice to show that for any bounded, invariant random variable W

$$\hat{E}^x W = 0 \quad \text{for all } x \in \Theta, \quad \Longrightarrow \quad \hat{P}^x \{W = 0\} = 1 \quad \text{for all } x \in \Theta. \tag{9.1.8}$$

Let $C = \cap_{i \le m} \left\{ \hat{X}_i = x_i \right\}$ be any cylinder event with positive $\hat{P}^x$-probability. Since W is $\mathcal{I}$-measurable, it satisfies $W = W \circ \sigma^m$, and so by the Markov property

$$\begin{aligned} \hat{E}^x(W\mathbf{1}_C) &= \hat{E}^x W(\hat{X}_0, \hat{X}_1, \cdots)\mathbf{1}_C \\ &= \hat{E}^x W(\hat{X}_m, \hat{X}_{m+1}, \cdots)\mathbf{1}_C \\ &= \hat{E}^x(W(\hat{X}_m, \hat{X}_{m+1}, \cdots) \,|\, C)\hat{P}^x(C) \\ &= \hat{E}^{x_m} W(\hat{X}_0, \hat{X}_1, \cdots)\hat{P}^x(C). \end{aligned}$$

But

$$\hat{E}^{x_m} W(\hat{X}_0, \hat{X}_1, \cdots) = 0,$$

by hypothesis; consequently, $\hat{E}^x(W\mathbf{1}_C) = 0$ for every cylinder event C. This implies that $\hat{P}^x \{W = 0\} = 1$, by the following elementary fact from integration theory. □

Lemma 9.1.6 *Let Z be a bounded, Borel-measurable random variable on Θ^∞. If $\hat{E}^x(Z\mathbf{1}_C) = 0$ for every cylinder event $C = \cap_{i=0}^m \left\{ \hat{X}_i = x_i \right\}$ then $\hat{P}^x \{Z = 0\} = 1$.*

Proof. Let $Z = Z_+ - Z_-$ be the decomposition of Z into its positive and negative parts. The lemma can be restated in terms of this decomposition as follows: *if $\hat{E}^x(Z_+\mathbf{1}_C) = \hat{E}^x(Z_+\mathbf{1}_C)$ for every cylinder set C, then $\hat{P}^x \{Z_+ = Z_-\} = 1$.*

The random variables Z_+, Z_- are the Radon-Nikodym derivatives of finite measures Q_+, Q_- on $\mathcal{B}_\infty$, defined by

$$\begin{aligned} Q_+(F) &= \hat{E}^x(Z_+\mathbf{1}_F) \quad \text{and} \\ Q_-(F) &= \hat{E}^x(Z_-\mathbf{1}_F) \quad \text{for all } F \in \mathcal{B}_\infty. \end{aligned}$$

(See Section A.8 of the Appendix; that Q_+ and Q_- are countably additive is a consequence of the Monotone Convergence Theorem.) The hypothesis asserts that these two measures agree on cylinder events. Since the Borel σ-algebra $\mathcal{B}_\infty$ is the smallest σ-algebra containing all cylinders, it follows that the measures Q_+ and Q_- coincide on $\mathcal{B}_\infty$. Thus,

$$\hat{E}^x(Z_+\mathbf{1}_F) = \hat{E}^x(Z_+\mathbf{1}_F) \quad \text{for every } F \in \mathcal{B}_\infty,$$

and so it follows that $Z_+ = Z_-$ almost surely. □

9.2 Absolute Continuity of Exit Measures

Definition 9.2.1 A Markov chain $(X_n)_{n\geq 0}$ (or its transition probability matrix p) is said to be *rooted* if there is a distinguished state x_*, called the *root*, from which every other state is accessible, that is, such that for every $y \in \Theta$ there exists a nonnegative integer n such that $p_n(x_*, y) > 0$. For a rooted Markov chain, write $\nu = \nu_{x_*}$.

If a Markov chain $\mathbf{X} = (X_n)_{n\geq 0}$ is irreducible, then it is obviously rooted; any state can be designated as the root. There are interesting Markov chains that are rooted but not irreducible — see, for instance, Example 9.3.4 in Section 9.3 below.

Proposition 9.2.2 *Assume that the Markov chain* $\mathbf{X}$ *is rooted with root state* x_* *and distinguished exit measure* $\nu = \nu_{x_*}$. *Then for every state* $x \in \Theta$ *the exit measure* ν_x *is absolutely continuous with respect to* ν, *i.e., for every event* $F \in \mathcal{I}$,

$$\nu(F) = 0 \implies \nu_x(F) = 0.$$

Furthermore, the likelihood ratio

$$L_x = \left(\frac{d\nu_x}{d\nu}\right)_{\mathcal{I}} \tag{9.2.1}$$

is bounded above by $1/\sup_{n\geq 0} p_n(x_*, x)$.

Proof. According to Proposition 9.1.4, the exit measures satisfy the mean value property (9.1.5). Iteration of this relation shows that for any $x \in \Theta$ and every integer $n \geq 1$,

$$\nu_x = \sum_{y\in\Theta} p_n(x, y)\nu_y.$$

It follows that for any two states x, y, any $n \in \mathbb{N}$, and every event $F \in \mathcal{I}$,

$$\nu_x(F) \geq p_n(x, y)\nu_y(F),$$

and so for any states x, y for which $\sup_{n\geq 0} p_n(x, y) > 0$ the measure ν_y must be absolutely continuous with respect to ν_x, with likelihood ratio no larger than $1/\sup_{n\geq 0} p_n(x, y) > 0$. □

Corollary 9.2.3 *For a rooted Markov chain, the vector space of bounded harmonic functions is linearly isomorphic to* $L^\infty(\Theta^\infty, \mathcal{I}, \nu)$, *by the mapping that sends any*

element $Y \in L^\infty(\Theta^\infty, \mathcal{I}, \nu)$ *to the harmonic function*

$$u(x) = \int Y L_x \, d\nu. \tag{9.2.2}$$

Proof. By Theorem 9.1.5, the equation (9.1.7) establishes a linear bijection between bounded harmonic functions u and equivalence classes of bounded, $\mathcal{I}$-measurable random variables Y, where Y_1, Y_2 are equivalent if and only if $\nu_x \{Y_1 \neq Y_2\} = 0$ for every state x. Proposition 9.2.2 implies that if the Markov chain is rooted, then each exit measure ν_x is absolutely continuous with respect to ν, and so Y_1, Y_2 are equivalent if and only if $\nu \{Y_1 \neq Y_2\} = 0$, that is, if Y_1 and Y_2 are equal as elements of $L^\infty(\Theta^\infty, \mathcal{I}, \nu)$. The identity (9.2.2) follows directly from equation (9.1.6) and the Radon-Nikodym theorem. □

Exercise 9.2.4 Show by example that for any integer $k \geq 1$ there is a rooted Markov chain for which the vector space of bounded harmonic functions has dimension k.

Definition 9.2.5 A Markov chain (or its transition probability kernel) has the *Liouville property* if it admits no nonconstant bounded harmonic functions.

Corollary 9.2.6 *For a rooted Markov chain the following properties are equivalent.*

(A) The Markov chain has the Liouville property.
(B) The exit measures ν_x are all equal.
(C) The invariant σ-algebra is trivial under ν, that is, there is no event $A \in \mathcal{I}$ such that $0 < \nu(A) < 1$.

Proof. That (B) implies (A) follows immediately from Theorem 9.1.5, since this implies that every bounded harmonic function u has an essentially unique representation $u(x) = E_{\nu_x} Y$ for some bounded, $\mathcal{I}$-measurable random variable Y. That (C) implies (B) follows from the absolute continuity of the measures ν_x with respect to ν, because if $\mathcal{I}$ is trivial then every $\mathcal{I}$-measurable random variable — in particular, every likelihood ratio L_x — must be essentially constant. But if $\nu \{L_x = c\} = 1$ for some $c \in \mathbb{R}$ then $c = 1$, since ν_x is a probability measure, and hence $\nu_x = \nu$.

To prove that (A) implies (C), suppose to the contrary that there is an event $A \in \mathcal{I}$ such that $0 < \nu(A) < 1$. Let $u(x) = E_{\nu_x} \mathbf{1}_A$ be the corresponding harmonic function. Clearly, $u(x_*) = E_\nu \mathbf{1}_A = \nu(A)$; but by (9.1.7), the limiting value $\lim_{n\to\infty} u(X_n) = \mathbf{1}_A \circ \mathbf{X}$ along almost any random walk path must be 0 or 1, so u cannot be a constant function. □

Proposition 9.2.7 *If the Markov chain* $\mathbf{X}$ *is rooted, then for ν-almost every* $\mathbf{y} \in \Theta^\infty$ *the function* $x \mapsto L_x(\mathbf{y})$ *is harmonic.*

Proof. This follows from the correspondence (9.2.2) by a standard argument. For any event $F \in \mathcal{I}$, the function $x \mapsto E_\nu(\mathbf{1}_F L_x)$ is harmonic, and so

$$E_\nu(\mathbf{1}_F L_x) = E_\nu\left(\mathbf{1}_F \sum_{y\in\Theta} p(x,y)L_y\right).$$

By an elementary theorem of integration theory, any two integrable random variables that have the same integrals on all measurable sets must be equal almost surely; hence, $L_x = \sum_{y\in\Theta} p(x,y)L_y$ almost surely. Since the set of states x is countable, the null events on which these identities fail are contained in a single null event; for every $\mathbf{y} \in \Theta^\infty$ not in this null event, the function $x \mapsto L_x(\mathbf{y})$ is harmonic. $\square$

9.3 Two Examples

Blackwell's Theorem 9.1.5 and its Corollary 9.2.3 for rooted Markov chains characterize the spaces of bounded harmonic functions by giving explicit isomorphisms with the Banach spaces $L^\infty(\Theta^\infty, \mathcal{I}, \nu)$. The measure space $(\Theta^\infty, \mathcal{I}, \nu)$ is a slippery object, however; the σ-algebra $\mathcal{I}$ need not even be countably generated. Thus, the problem of finding integral representations for bounded harmonic functions does not end with Blackwell's Theorem. In Chapter 12, we will undertake a detailed study of *Furstenberg-Poisson boundaries* — topological measure spaces that provide representations of bounded harmonic functions — for random walks on finitely generated groups. In this section, we will look at two concrete and relatively elementary examples that illustrate some of the main issues.

Suppose that the state space Θ of a Markov chain $(X_n)_{n\geq 0}$ is embedded in a metrizable topological space $\mathcal{Y}$ in such a way that $\lim_{n\to\infty} X_n := X_\infty \in \mathcal{Y}$ exists P^x-almost surely for every $x \in \Theta$. (Such embeddings always exist: in the one-point compactification of Θ, for example, every sequence that eventually leaves every finite subset of Θ must converge to the (sole) boundary point.) The existence and distribution of the limit depend only on the *law* $\hat{P}^x$ of the Markov chain; thus, almost sure convergence holds if and only if

$$\hat{X}_\infty := \lim_{n\to\infty} \hat{X}_n \tag{9.3.1}$$

exists almost surely $\hat{P}^x$ for every $x \in \Theta$, where $\hat{X}_n : \Theta^\infty \to \Theta$ is the nth coordinate random variable. The limit $\hat{X}_\infty$ is (after modification on an event of probability zero — cf. Exercise 9.1.2) measurable with respect to the invariant σ-algebra $\mathcal{I}$. Therefore, Proposition 9.1.4 implies that for any bounded, Borel measurable function $f : \mathcal{Y} \to \mathbb{R}$ the function $u : \Theta \to \mathbb{R}$ defined by

$$u(x) := \hat{E}^x f(\hat{X}_\infty) = E_{\nu_x} f(\hat{X}_\infty) \tag{9.3.2}$$

is bounded and harmonic. When is it the case that *all* bounded harmonic functions arise in this way?

Example 9.3.1 (Nearest Neighbor Markov Chains on a Homogeneous Tree) Let $\mathbb{T} = \mathbb{T}_d$ be the infinite homogeneous tree of degree $d \geq 2$, and let Θ be its vertex set. A Markov chain on the set Θ is a *nearest neighbor Markov chain* if its transition probability matrix p is supported by the edge set of $\mathbb{T}$, that is, if $p(x, y) > 0$ only if the vertices x, y are nearest neighbors. If $(X_n)_{n \geq 0}$ is a *transient*, irreducible, nearest neighbor Markov chain on $\mathbb{T}$ then for every vertex x,

$$\hat{P}^x \left\{ \hat{X}_\infty := \lim_{n \to \infty} \hat{X}_n \text{ exists} \right\} = 1. \tag{9.3.3}$$

(See Exercise 1.6.6 for the special case where the Markov chain is a random walk. The general case is no different.) The limit $\hat{X}_\infty$ is an $\mathcal{I}$-measurable random variable valued in the space $\partial \mathbb{T}$ of ends of the tree (cf. Exercise 9.1.2).

Exercise 9.3.2 This exercise outlines a proof that every bounded, harmonic function u for a transient, irreducible, nearest neighbor Markov chain on the tree $\Theta = \mathbb{T} = \mathbb{T}_d$ has the form $u(x) = \hat{E}^x f(\hat{X}_\infty)$ for some bounded, Borel measurable function $f : \partial \mathbb{T} \to \mathbb{R}$.

(A) Let u be a bounded, harmonic function on Θ. Show that for ν-almost every $\omega = a_1 a_2 \cdots \in \partial \mathbb{T}$,

$$\lim_{n \to \infty} u(a_1 a_2 \cdots a_n) := f(\omega) \quad \text{exists.}$$

Hint: $\lim_{n \to \infty} u(\hat{X}_n)$ exists P^x-almost surely for any $x \in \Theta$, by the Martingale Convergence Theorem. What you must show is that the limit depends only on $\hat{X}_\infty$.

(B) Show that f can be extended to the entire space $\partial \mathbb{T}$ in such a way that it is Borel measurable and bounded, and conclude that for this extension

$$\lim_{n \to \infty} u(\hat{X}_n) = f(\hat{X}_\infty) \quad \hat{P}^x - \text{almost surely,}$$

and therefore $u(x) = \hat{E}^x f(\hat{X}_\infty)$.

Let's return to the general problem of integral representations of bounded harmonic functions. Assume once again that $(X_n)_{n \geq 0}$ is a transient Markov chain whose state space Θ is embedded in a metrizable space $\mathcal{Y}$ and that $\lim_{n \to \infty} X_n := X_\infty \in \mathcal{Y}$ exists almost surely. When does every bounded harmonic function u have a representation (9.3.2)? By Theorem 9.1.5, a necessary and sufficient condition is that

$$L^\infty(\Theta^\infty, \mathcal{G}, \nu_x) = L^\infty(\Theta^\infty, \mathcal{I}, \nu_x) \quad \text{for every} \quad x \in \Theta, \tag{9.3.4}$$

where $\mathcal{G} = \sigma(\hat{X}_\infty) \subset \mathcal{I}$ is the σ-algebra generated by the random variable $\hat{X}_\infty$. Since $\mathcal{G} \subset \mathcal{I}$, the equality (9.3.4) holds if and only if the invariant σ-algebra $\mathcal{I}$ is a *null extension* of $\mathcal{G}$, that is, if every event $F \in \mathcal{I}$ differs from an event $F' \in \mathcal{G}$ by an event $F \Delta F'$ of ν_x-probability zero. If the Markov chain is rooted, with root state x_*, then the exit measures ν_x are all absolutely continuous with respect to $\nu = \nu_{x_*}$, so condition (9.3.4) is equivalent to

$$L^\infty(\Theta^\infty, \mathcal{G}, \nu) = L^\infty(\Theta^\infty, \mathcal{I}, \nu). \tag{9.3.5}$$

Proposition 9.3.3 *Assume that $(X_n)_{n \geq 0}$ is a rooted Markov chain on the state space Θ with distinguished exit measure ν and invariant σ-algebra $\mathcal{I}$. For a σ-algebra $\mathcal{G} \subset \mathcal{I}$, the equality (9.3.5) holds if and only if the likelihood ratios $L_x := (d\nu_x/d\nu)$ all have $\mathcal{G}$-measurable versions.*

Proof. The forward implication is clear, because the hypothesis $L^\infty(\Theta^\infty, \mathcal{G}, \nu) = L^\infty(\Theta^\infty, \mathcal{I}, \nu)$ implies that each likelihood ratio L_x, which by the Radon-Nikodym theorem is $\mathcal{I}$-measurable, is almost surely equal to a $\mathcal{G}$-measurable random variable.

Conversely, suppose that for each $x \in \Theta$ there is a version L_x of the likelihood ratio $(d\nu_x/d\nu)_{\mathcal{I}}$ that is $\mathcal{G}$-measurable. Let $A \in \mathcal{I}$ be an invariant event and let $u : \Gamma \to \mathbb{R}$ be the bounded harmonic function $u(x) = E_\nu(\mathbf{1}_A L_x)$. Since L_x is $\mathcal{G}$-measurable, the defining property of conditional expectation (see Section A.9 of the Appendix) implies that for every $x \in \Theta$,

$$u(x) = E_\nu(\mathbf{1}_A L_x) = E_\nu(E_\nu(\mathbf{1}_A \,|\, \mathcal{G}) L_x).$$

Thus, the harmonic function u has two representations (9.1.6), one with the $\mathcal{I}$-measurable random variable $\mathbf{1}_A$ and the other with the $\mathcal{G}$-measurable random variable $E_\nu(\mathbf{1}_A \,|\, \mathcal{G})$. Since these two functions are both bounded and $\mathcal{I}$-measurable, Theorem 9.1.5 implies that they must be almost surely equal. This in turn implies that the event A differs from an event $A' \in \mathcal{G}$ by an event of probability 0. □

One way to verify the hypothesis of Proposition 9.3.3 is to exhibit a sequence of σ-algebras $\mathcal{H}_n$ all containing $\mathcal{I}$ such that the likelihood ratios

$$Z_n := \left(\frac{d\hat{P}^x}{d\hat{P}} \right)_{\mathcal{H}_n}$$

converge in L^1 norm (relative to $\hat{P}$) to a $\mathcal{G}$-measurable random variable L_x. It then follows that L_x is a version of $(d\nu_x/d\nu)$, because for every event $F \in \mathcal{I} \subset \mathcal{H}_n$,

$$E_\nu(L_x \mathbf{1}_F) = \lim_{n \to \infty} \hat{E}(Z_n \mathbf{1}_F) = \hat{P}^x(F) = \nu_x(F).$$

Example 9.3.4 (Pólya's Urn) † This is a rooted — but not irreducible — Markov chain $(X_n)_{n\geq 0} = (R_n, W_n)_{n\geq 0}$ on the positive quadrant $\Theta = \mathbb{N} \times \mathbb{N}$ of the two-dimensional integer lattice with root state (1, 1) and transition probabilities

$$p((r, w), (r+1, w) = r/(r+w), \tag{9.3.6}$$

$$p((r, w), (r, w+1)) = w/(r+w). \tag{9.3.7}$$

The coordinates R_n and W_n represent the numbers of red and white balls in an "urn" at time n. The transition probabilities translate as follows: on the $(n+1)$th step, one of the $R_n + W_n$ balls in the urn is selected at random and then returned to the urn along with a new ball of the same color. (Thus, the total number $R_n + W_n$ of balls after n steps is $n + R_0 + W_0$.)

Exercise 9.3.5 †

(A) Show that under $P^{(1,1)}$ the distribution of R_n is the uniform distribution on $\{1, 2, \cdots, n+1\}$.
(B) Show that the function $h(r, w) = r/(r+w)$ is harmonic.
(C) Show that under $P^{(1,1)}$ the almost sure limit $U := \lim_{n\to\infty} h(X_n)$ has the uniform distribution on the unit interval [0, 1]. In Exercise 9.3.6 you will show that every bounded harmonic function $u : \Theta \to \mathbb{R}$ has the form $u(x) = E^x f(U)$ for some bounded, Borel measurable $f : [0, 1] \to \mathbb{R}$.
(D) Show that there is a metrizable topology on $(\mathbb{N} \times \mathbb{N}) \cup [0, 1]$ such that (9.3.1) holds.

Exercise 9.3.6 † This exercise outlines a proof that for Pólya's urn every bounded harmonic function has a representation $u(x) = E(f(U)L_x)$ where U is the limit random variable of Exercise 9.3.5 and L_x is the likelihood ratio $d\nu_x/d\nu$. For ease of notation, assume that $\hat{P}^x = P^x$, $\hat{X}_n = X_n$, etc. For each $m \in \mathbb{N}$, let B_m be the set of all points (r, w) in the state space $\Theta = \mathbb{N} \times \mathbb{N}$ such that $r + w < m$. Define random variables

$$\tau_m := \min\{n \geq 0 \,:\, X_n \notin B_m\}$$

and let $\mathcal{H}_m$ be the σ-algebra generated by the random variables $X_{\tau_m}, X_{\tau_m+1}, X_{\tau_m+2}, \cdots$. Fix a state $x = (r, w) \in \Theta$, and set $m = r + w$. Write $P = P^{(1,1)}$.

(A) Verify that the σ-algebras $\mathcal{H}_m$ all contain $\mathcal{I}$.
(B) Show that for any integers $r', w' \geq 0$ such that $r' + w' = m'$

$$P^x\left\{(R_{\tau_{m+m'}}, W_{\tau_{m+m'}}) = (r+r', w+w')\right\} = \binom{m'}{r'} \frac{\prod_{i=0}^{r'-1}(r+i)\prod_{j=0}^{w'-1}(w+j)}{\prod_{k=0}^{m'-1}(m+k)}.$$

(C) Use (B) to show that on $\{(R_{\tau_{m+m'}}, W_{\tau_{m+m'}}) = (r + r', w + w')\}$,

$$\left(\frac{dP^x}{dP}\right)_{\mathcal{H}_{m+m'}} = \prod_{i=0}^{r-1}\left(\frac{r'+i}{1+i}\right)\prod_{j=0}^{w-1}\left(\frac{w'+j}{1+j}\right)\Big/\prod_{k=0}^{m-1}\left(\frac{m'+k}{2+k}\right)$$

provided $r', w' \geq 0$ are such that $r' + w' = m'$, and

$$\left(\frac{dP^{x_*}}{dP}\right)_{\mathcal{H}_{m+m'}} = 0 \quad \text{otherwise.}$$

(D) Show that

$$\lim_{m'\to\infty}\left(\frac{dP^x}{dP}\right)_{\mathcal{H}_{m+m'}} = (m-1)\binom{m-2}{r-1}U^{r-1}(1-U)^{w-1} \tag{9.3.8}$$

both almost surely and in L^1 norm (with respect to P).

(E) Conclude that the right side of (9.3.8) is a version of $(d\nu_x/d\nu)_{\mathcal{I}}$, and deduce from this that severy bounded harmonic function is of the form $u(x) = E^x f(U)$ for some bounded function $f : [0, 1] \to \mathbb{R}$.

9.4 The Tail σ-Algebra $\mathcal{T}$

A transition probability matrix p has the same harmonic functions as the *lazy* transition probability matrix $p'(x, y) := (p(x, y) + \delta(x, y))/2$. For a lazy Markov chain the invariant σ-algebra has as a null extension the *tail σ-algebra*, as we will show in Proposition 9.4.3. This is defined as follows.

Definition 9.4.1 Let $\hat{X}_n : \Theta^\infty \to \Theta$ be the nth coordinate projection. The *tail σ-algebra* $\mathcal{T}$ on Θ^∞ is the intersection

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(\hat{X}_n, \hat{X}_{n+1}, \hat{X}_{n+2}, \cdots) \tag{9.4.1}$$

where $\sigma(\hat{X}_n, \hat{X}_{n+1}, \hat{X}_{n+2}, \cdots)$ is the σ-algebra generated by the random variables $\hat{X}_n, \hat{X}_{n+1}, \hat{X}_{n+2}, \cdots$. Equivalently, $\mathcal{T}$ is the σ-algebra of all events $F \in \mathcal{B}_\infty$ such that for any two sequences $\mathbf{y} = (y_0, y_1, \cdots)$ and $\mathbf{y}' = (y_0', y_1', \cdots)$ that agree in all but finitely many entries,

$$\mathbf{y} \in F \iff \mathbf{y}' \in F. \tag{9.4.2}$$

Exercise 9.4.2 (A) Show that $\mathcal{I} \subset \mathcal{T}$. (B) Show that if $|\Theta| \geq 2$ then $\mathcal{T}$ is strictly larger than $\mathcal{I}$.

HINT: Look for a tail event that has some built in periodicity.

Proposition 9.4.3 *For a lazy Markov chain $(\hat{X}_n)_{n\geq 0}$ the tail σ-algebra is a* null extension *of the invariant σ-algebra, that is, for any $\mathcal{T}$-measurable random variable $Y : \Theta^\infty \to \mathbb{R}$ there exists an $\mathcal{I}$-measurable random variable $Y' : \Theta^\infty \to \mathbb{R}$ such that for every $x \in \Theta$,*

$$\hat{P}^x \left\{ Y = Y' \right\} = 1. \tag{9.4.3}$$

Consequently, if the Markov chain $(X_n)_{n\geq 0}$ is lazy and rooted, with root state x_*, then the space of bounded harmonic functions is linearly isomorphic to $L^\infty(\Theta^\infty, \mathcal{T}, \hat{P}^{x_*})$.

We will deduce this from the following *shift-coupling* construction.

Lemma 9.4.4 *If the transition probability matrix p satisfies $p(x, x) \geq r > 0$ for every state x, then for every state x there is a probability space $(\Omega, \mathcal{F}, P)$ on which are defined two versions X'_n and X''_n of the Markov chain, both with initial point $X'_0 = X''_0 = x$ and transition probability matrix p, such that with probability* 1,

$$X''_n = X'_{n+1} \quad \textit{eventually.} \tag{9.4.4}$$

Similarly, for any $m = 0, 1, 2, \cdots$, there exist versions X'_n and X''_n of the Markov chain, both with initial point $X'_0 = X''_0 = x$, such that with probability 1,

$$\begin{aligned} X''_n &= X'_n \quad \textit{for all } n \leq m, \quad \textit{and} \\ X''_n &= X'_{n+1} \quad \textit{eventually.} \end{aligned} \tag{9.4.5}$$

Observe that (9.4.4) implies

$$\lim_{n\to\infty} \sum_{y\in\Theta} |p_n(x, y) - p_{n+1}(x, y)| = 0. \tag{9.4.6}$$

Proof. The transition probability matrix p of a lazy Markov chain with holding probability $r > 0$ has the form $p = (1 - r)q + r\delta$, where δ is the Kronecker delta and q is another transition probability matrix. A Markov chain with transition probability matrix p evolves as follows: at each time $n = 1, 2, \cdots$ the walker either stays put or makes a step using the transition probability matrix q, depending on the result of an independent r-coin toss. Thus, versions of the Markov chain can be constructed on any probability space $(\Omega, \mathcal{F}, P^x)$ that supports a Markov chain $(Y_n)_{n\geq 0}$ with transition probability matrix q and initial state $Y_0 = x$ and independent sequences $\{U_n\}_{n\geq 1}$and $\{V_n\}_{n\geq 1}$ of i.i.d. Bernoulli-r random variables. (The standard

Lebesgue space $([0,1], \mathcal{B}_{[0,1]}, \text{Leb})$ can always be used – cf. Section A.4 in the Appendix.) Set

$$S_n^U = \sum_{i=1}^n U_i \quad \text{and} \quad S_n^V = \sum_{i=1}^n V_i;$$

then the sequences

$$X'_n = Y_{S_n^U} \quad \text{and}$$
$$X^*_n = Y_{S_n^V}$$

are both Markov chains with transition probability matrix p and initial state x. Now the sequence $S_n^U - S_n^V$ is an aperiodic, symmetric, nearest-neighbor random walk on $\mathbb{Z}$, so by Pólya's recurrence theorem it visits every integer, with probability 1. Let T be the least positive integer n such that $S_n^V = S_n^U + 1$, and define

$$X''_n = X^*_n \quad \text{for all } n \le T,$$
$$= X'_{n+1} \quad \text{for all } n > T$$

The sequence $(X''_n)_{n \ge 0}$ is a Markov chain with transition probability matrix p (Exercise: Check this.), and by construction it has the property (9.4.4).

A simple modification of this construction produces a version of the Markov chain such that (9.4.5) holds: just toss the same coin for both versions of the random walk until time m, that is, replace the definition of S_n^V above by

$$S_n^V = S_{n\wedge m}^U + \sum_{i=m+1}^n V_i.$$

$\square$

Proof of Proposition 9.4.3 Fix $x \in \Theta$, and let $Y : \Theta^\infty \to \mathbb{R}$ be a bounded, $\mathcal{T}$-measurable random variable. We will prove that

$$\hat{P}^x \{Y = Y \circ \sigma\} = 1 \quad \text{for every} \quad x \in \Theta, \tag{9.4.7}$$

where σ denotes the forward shift mapping (2.1.3) and $\hat{P}^x$ is the law of the Markov chain with transition probability matrix p and initial point x. By the Markov Property (cf. Proposition 6.1.8) and the countability of the state space Θ, this will imply that for every $x \in \Theta$,

$$\hat{P}^x(G) = 1 \quad \text{where} \quad G := \left\{Y = Y \circ \sigma^n \text{ for every } n \in \mathbb{N}\right\}.$$

(Exercise: Explain this.) Consequently, if $Y' : \Theta^\infty \to \mathbb{R}$ is defined by

$$Y'(\mathbf{x}) = Y(\mathbf{x}) \quad \text{for all } \mathbf{x} \in G,$$
$$Y'(\mathbf{x}) = Y(\sigma^n \mathbf{x}) \quad \text{if } \sigma^n \mathbf{x} \in G, \quad \text{and}$$
$$Y'(\mathbf{x}) = 0 \quad \text{if } \sigma^n \mathbf{x} \in G^c \quad \text{for all } n \in \mathbb{Z}_+;$$

then Y' satisfies $Y' = Y' \circ \sigma$ everywhere on Θ^∞, and hence is $\mathcal{I}$-measurable. Clearly, $\hat{P}^x \left\{ Y = Y' \right\} = 1$, since $Y = Y'$ on the event G.

To prove the equality (9.4.7) it suffices, by Lemma 9.1.6, to prove that for every cylinder event $\hat{C} = \cap_{i=0}^m \left\{ \hat{X}_i = x_i \right\}$,

$$\hat{E}^x Y \mathbf{1}_{\hat{C}} = \hat{E}^x (Y \circ \sigma) \mathbf{1}_{\hat{C}}. \tag{9.4.8}$$

By Lemma 9.4.4, there is a probability space $(\Omega, \mathcal{F}, P)$ that supports versions $(X'_n)_{n \geq 0}$ and $(X''_n)_{n \geq 0}$ of the Markov chain, both with initial point $X'_0 = X''_0 = x$, that satisfy the shift-coupling condition (9.4.5). Because Y is $\mathcal{T}$-measurable, it does not depend on any finite set of coordinates; therefore, since $X'_{n+1} = X''_n$ eventually with P-probability 1,

$$Y(X''_0, X''_1, X''_2, \cdots) = Y(X'_1, X'_2, X'_3, \cdots) \quad P\text{−almost surely.}$$

Furthermore, since the trajectories of X''_n and X'_n coincide until time $n = m$, for any cylinder set

$$C = \bigcap_{i=0}^m \left\{ X'_i = x_i \right\} = \bigcap_{i=0}^m \left\{ X''_i = x_i \right\}$$

involving only the first $m + 1$ coordinates,

$$Y(X''_0, X''_1, X''_2, \cdots) \mathbf{1}_C = Y(X'_0, X'_1, X'_2, \cdots) \mathbf{1}_C \quad P\text{−almost surely.}$$

Thus, for any $x \in \Theta$,

$$\begin{aligned} E_P Y(X'_0, X'_1, X'_2, \cdots) \mathbf{1}_C &= E_P Y(X''_0, X''_1, X''_2, \cdots) \mathbf{1}_C \\ &= E_P Y(X'_1, X'_2, X'_3, \cdots) \mathbf{1}_C, \end{aligned}$$

and so (9.4.8) follows. □

9.5 Weak Ergodicity and the Liouville Property

Some Markov chains (cf. Exercise 9.2.4) have an abundance of nonconstant bounded harmonic functions, but others have none. Markov chains with no nontriv-

ial bounded harmonic functions are said to have the *Liouville property*. According to Corollary 9.2.6, a rooted Markov chain has the Liouville property if and only if the invariant σ-algebra $\mathcal{I}$ is trivial for the exit measure ν. This leaves us with the not-always-easy problem of determining whether or not there are invariant events with probabilities between 0 and 1. In this section and the next we will formulate several alternative sufficient conditions.

If $h : \Theta \to \mathbb{R}$ is a bounded harmonic function for the Markov chain $(X_n)_{n\geq 0}$ then for any state x, the sequence $(h(X_n))_{n\geq 0}$ is a martingale under P^x. Hence, for any $n \in \mathbb{N}$,

$$h(x) = E^x h(X_n). \tag{9.5.1}$$

This suggests that if for large n the distributions of X_n under the probability measures P^x and P^z do not differ much, then $h(x) \approx h(z)$. The right way to measure the difference between two probability distributions on a discrete set is by *total variation distance.*

Definition 9.5.1 The total variation distance between two probability distributions λ_1 and λ_2 on a countable set Θ is defined by

$$\|\lambda_1 - \lambda_2\|_{TV} := \sum_{y\in\Theta} |\lambda_1(y) - \lambda_2(y)|. \tag{9.5.2}$$

A Markov chain $(X_n)_{n\geq 0}$ on a countable set Θ (or its transition probability kernel) is said to be *weakly ergodic* if for any two states x, z,

$$\begin{aligned} \lim_{n\to\infty} \left\| P^x\{X_n = \cdot\} - P^z\{X_n = \cdot\} \right\|_{TV} = 0 \quad &\Longleftrightarrow \\ \lim_{n\to\infty} \sum_{y\in\Theta} |p_n(x, y) - p_n(z, y)| = 0. \end{aligned} \tag{9.5.3}$$

Exercise 9.5.2 Show that for any two states x, z the total variation distance (9.5.3) is nonincreasing in n.

Exercise 9.5.3 Show that for any *symmetric*, weakly ergodic Markov kernel p on an infinite state space Θ

$$\lim_{n\to\infty} p_n(x, y) = 0 \quad \text{for all } \ x, y \in \Theta. \tag{9.5.4}$$

Proposition 9.5.4 *If a Markov chain is weakly ergodic then it has the Liouville property.*

Proof. Let $(X_n)_{n\geq 0}$ be a weakly ergodic Markov chain with transition probabilities $p(\cdot, \cdot)$. If h is a bounded, harmonic function then by equation (9.5.1), for any two states x, z and any $n \geq 1$,

$$\begin{aligned}|h(x) - h(z)| &= \Big|\sum_{y\in\Theta} h(y)(P^x\{X_n = y\} - P^z\{X_n = y\})\Big| \\ &\leq \|h\|_\infty \left\| P^x\{X_n = \cdot\} - P^z\{X_n = \cdot\}\right\|_{TV};\end{aligned}$$

weak ergodicity implies that this converges to 0 as $n \to \infty$. □

Thus, weak ergodicity is a sufficient condition for the Liouville property. It is not a *necessary* condition, however. For example, the simple random walk on $\mathbb{Z}$ has the Liouville property, because it is recurrent (cf. Exercise 8.0.2), but it is not weakly ergodic. The obstruction is *periodicity*: if two states $x, z \in \mathbb{Z}$ have oppositive parity, then the distributions $p_n(x, \cdot)$ and $p_n(z, \cdot)$ have disjoint supports, so their total variation distance is 2. For *lazy* Markov chains (those whose transition probability kernels satisfy $p(x, x) \geq r > 0$ for some constant $r > 0$ not depending on x) this problem disappears.

Theorem 9.5.5 *A lazy, irreducible Markov chain has the Liouville property if and only if it is weakly ergodic.*

The proof of the forward implication will be given in Section 9.7.

9.6 Coupling

There is a useful indirect approach to establishing weak ergodicity that does not require direct estimates on transition probabilities (but can, in some cases, provide sharp estimates for total variation distances). This is the *coupling method*.

Definition 9.6.1 Let p be a Markov kernel on a finite or countable state space Θ. Given states $x, z \in \Theta$, a *coupling* of Markov chains with transition probabilities $p(\cdot, \cdot)$ and initial points x, z is a sequence $(X_n^x, X_n^z)_{n\geq 0}$ of ordered pairs of Θ-valued random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$ such that

(a) for each of $i = x$ and $i = z$ the sequence $(X_n^i)_{n\geq 0}$ is a Markov chain with Markov kernel p and initial point $X_0^i = i$; and
(b) the sequences X_n^x and X_n^z eventually coincide, i.e.,

$$P\left\{X_n^x = X_n^z \text{ eventually}\right\} = 1. \tag{9.6.1}$$

The Markov kernel p (or the Markov chain governed by the Markov kernel p) is said to have the *coupling property* if for any two initial points $x, z \in \Theta$ there is a coupling.

Exercise 9.6.2 Show that if a Markov chain has the coupling property then it is weakly ergodic.

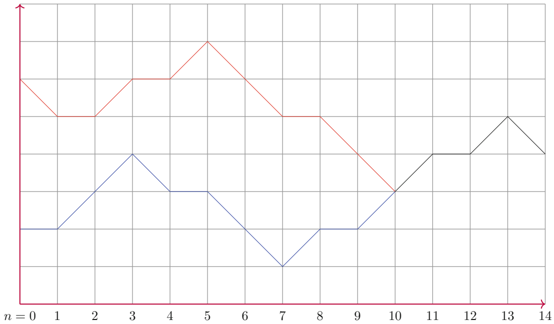


Fig. 9.1 Coupling Lazy Simple Random Walks on $\mathbb{Z}$

HINT: Show that the total variation distance $\|p_n(x,\cdot)-p_n(z,\cdot)\|_{TV}$ is bounded by $2P\left\{X_n^x \neq X_n^z\right\}$.

Together with Proposition 9.5.4, Exercise 9.6.2 shows that any Markov chain with the coupling property has the Liouville property. For lazy Markov chains the converse is also true, but as we will have no further use for it, we omit the proof, which can be found in [58], (or, for a simpler proof, [108]).

Example 9.6.3 Lazy simple random walk on the integers $\mathbb{Z}$ has the coupling property. This is a consequence of Pólya's recurrence theorem for the simple random walk on $\mathbb{Z}$. To construct couplings, let $(\xi_n^x)_{n\in\mathbb{N},x\in\mathbb{Z}}$ be independent, identically distributed random variables indexed by $x \in \mathbb{Z}$ and $n \in \mathbb{N}$, with common distribution

$$P\left\{\xi_n^x = +1\right\} = P\left\{\xi_n^x = -1\right\} = (1-r)/2 \quad \text{and} \quad P\left\{\xi_n^x = 0\right\} = r > 0 \tag{9.6.2}$$

(Such a family can be defined on Lebesgue space: see Section A.4 of the Appendix.) For each $x \in \mathbb{Z}$ define

$$S_n^x = x + \sum_{i=1}^{n} \xi_i^x \quad \text{and} \quad \tau_x = \min\left\{n \geq 0 \,:\, S_n^x = S_n^0\right\}.$$

See Figure 9.1 for an example. Pólya's theorem implies that each of the random variables τ_x is finite, with probability one, because the sequence $(S_n^x - S_n^0)_{n\geq 0}$ of differences is a symmetric, irreducible random walk on $\mathbb{Z}$ with finitely supported sttep distribution. Now define

$$\begin{aligned} X_n^x &= S_n^x \quad \text{for } n \leq \tau_x, \quad \text{and} \\ &= S_n^0 \quad \text{for } n \geq \tau_x. \end{aligned}$$

Exercise 9.6.4 Verify that the sequence $(X_n^x)_{n\geq 0}$ is a random walk with initial point x and transition probabilities (9.6.2).

Exercise 9.6.5 Let $(X_n)_{n\geq 0}$ be the Markov chain on the first quadrant $\mathbb{Z}_+^2$ of the two-dimensional integer lattice with transition probabilities

$$\begin{aligned} p((x,y),(x+1,y)) &= p, \\ p((x,y),(x,y+1)) &= 1-p. \end{aligned} \tag{9.6.3}$$

Show that this Markov chain has the Liouville property.

HINT: Use a coupling argument to show that for any two states (x, y) and (x', y') such that $x + y = x' + y'$

$$\left\| p_n((x,y),\cdot) - p_n((x',y'),\cdot) \right\|_{TV} = 0.$$

Exercise 9.6.6 Show that lazy simple random walks on $\mathbb{Z}^d$ have the coupling property, for any dimension d.

HINT: Work on one coordinate at a time.

In fact, neither symmetry nor finite support of the step distribution is needed for the coupling property in $\mathbb{Z}^d$: all that is needed is irreducibility and aperiodicity. We won't prove this in full generality, since our main interest is the Liouville property; instead, we will use a coupling argument to prove the following proposition.

Proposition 9.6.7 *Every lazy, irreducible random walk on $\mathbb{Z}^d$ is weakly ergodic.*

Proof for $d = 1$**.** We will only prove the proposition for dimension $d = 1$; the higher dimensional cases can be proved by combining the ideas of this proof with the "one coordinate at a time" method of Exercise 9.6.6. To prove the result in the case $d = 1$, it suffices to prove it for random walks whose step distributions satisfy the stronger hypothesis that

$$\min(\mu(0), \mu(1)) > 0. \tag{9.6.4}$$

For if the step distribution μ did not satisfy (9.6.4) but did satisfy aperiodicity and irreducibility, then for some $k \geq 2$ the convolution power μ^{*k} would satisfy (9.6.4). If the random walk with step distribution μ^{*k} is weakly ergodic, then the random walk with step distribution μ is also weakly ergodic, by Exercise 9.5.2.

Suppose, then, that the step distribution μ satisfies (9.6.4); we will exhibit a coupling of random walks with step distribution μ and initial points 0 and $x \neq 0$. The idea, due to D. Ornstein, is to let the two random walkers make independent steps provided both make small steps, but to make the same step if one or the other random walker wants to make a big step. In detail: Let $(\xi_n, \xi_n')_{n\geq 1}$ be i.i.d. random variables with common distribution μ, and define

$$\begin{aligned} \xi_n'' &= \xi_n' \quad \text{if both } \xi_n, \xi_n' \in \{0,1\}; \quad \text{and} \\ \xi_n'' &= \xi_n \quad \text{otherwise.} \end{aligned}$$

The random variables $\xi''_1, \xi''_2, \cdots$ are i.i.d., because the rule for determining ξ''_j depends only on the pair (ξ_j, ξ'_j), and these are independent of the random vectors $(\xi_i, \xi'_i)_{i \neq j}$. Moreover, the distribution of ξ''_j is μ. (Exercise: Verify this.)

Fix an integer $x \neq 0$, and for each $n \geq 0$ define

$$S_n^x = x + \sum_{i=1}^{n} \xi''_i, \quad S_n^0 = 0 + \sum_{i=1}^{n} \xi_i, \quad \text{and}$$

$$\tau_x = \min \left\{ n \geq 1 \, : \, S_n^x = S_n^0 \right\}.$$

Pólya's theorem implies that $\tau_x < \infty$ with probability one, because the increments Y_n of the sequence $S_n^x - S_n^0$ are i.i.d. with distribution

$$P\{Y_n = +1\} = P\{Y_n = -1\} = \mu(0)\mu(1) \quad \text{and} \quad P\{Y_n = 0\} = 1 - 2\mu(0)\mu(1).$$

Now set

$$\begin{aligned} X_n^0 &= S_n^0, \\ X_n^x &= S_n^x \quad \text{if } n \leq \tau_x, \\ &= S_n^0 \quad \text{if } n \geq \tau_x; \end{aligned}$$

this is a coupling of random walks with initial points $0, x$, respectively. A similar construction can be given for arbitrary pairs x, x' of initial points. □

Exercise 9.6.8 Let Γ be a finitely generated group that has a finite index subgroup isomorphic to $\mathbb{Z}^d$. Show that any lazy, irreducible random walk on Γ whose step distribution has finite support is weakly ergodic.

HINT: Look back at Section 7.4.

Exercise 9.6.9 Let Γ be a finitely generated group for which there exists a surjective homomorphism $\varphi : \Gamma \to \mathbb{Z}^d$ with finite kernel $K := \ker(\varphi)$. Show that Γ has a finite-index subgroup that is isomorphic to $\mathbb{Z}^d$.

HINT: Let $g_1, g_2, \cdots, g_d \in \Gamma$ be elements such that $\varphi(g_i) = e_i$, where $e_1, e_2, \cdots, e_d$ are the standard unit vectors in $\mathbb{Z}^d$. (A) Show that for each i there is an automorphism $\psi_i : K \to K$ such that $g_i k = \psi_i(k) g_i$ for every $k \in K$. (B) Now use the hypothesis that $|K| < \infty$ to conclude that there is an integer $m \geq 1$ such that $g_i^m k = k g_i^m$ for every $k \in K$ and each index $i \in [d]$. (C) Show that for each pair i, j there is an element $k_{i,j} \in K$ such that $g_i^m g_j^m = k_{i,j} g_j^m g_i^m$. (D) Conclude that there is an integer $n \geq 1$ such that $g_i^n g_j^n = g_j^n g_i^n$ for every pair i, j.

9.7 Tail Triviality Implies Weak Ergodicity

Theorem 9.7.1 *A lazy, irreducible Markov chain has the Liouville property if and only if it is weakly ergodic.*

Proof. We have already proved the easy direction — that weak ergodicity implies the Liouville property — so only the forward implication remains. If a Markov chain $(X_n)_{n\geq 0}$ has the Liouville property then by Theorem 9.1.5, every event in its invariant σ-algebra has ν_x-probability 0 or 1, for every state x. Hence, by Proposition 9.4.3, if the Markov chain is also lazy then every event in its tail σ-algebra must also have ν_x-probability 0 or 1. Thus, our task is to show that for a lazy Markov chain tail triviality implies weak ergodicity.

We may assume that the underlying probability space is $(\Theta^\infty, \mathcal{F}_\infty, \hat{P}^x)$, where $\hat{P}^x$ is the law of the Markov chain with initial point x, and that the random variables $\hat{X}_n : \Theta^\infty \to \Theta$ are the coordinate projections. For ease of notation we will use the abbreviations $P^x = \hat{P}^x$ and $X_n = \hat{X}_n$. For each integer $n \geq 0$, let $\mathcal{G}_n = \sigma(X_n, X_{n+1}, \cdots)$. These σ-algebras are obviously nested, that is,

$$\mathcal{G}_0 \supset \mathcal{G}_1 \supset \mathcal{G}_2 \supset \cdots,$$

so they constitute a reverse filtration (cf. Definition 8.5.1) whose intersection is the tail σ-algebra $\mathcal{T}$. Fix two states $x, z \in \Theta$ for which $p_1(x, z) > 0$, and define a probability measure $Q^{x,z}$ on $(\Theta^\infty, \mathcal{G}_0)$ by

$$Q^{x,z}(F) = \frac{P^x(F \cap \{X_1 = z\})}{P^x(X_1 = z)} = \frac{P^x(F \cap \{X_1 = z\})}{p_1(x, z)}.$$

Clearly, the measure $Q^{x,z}$ is absolutely continuous with respect to P^x on $\mathcal{G}_0$, with Radon-Nikodym derivative bounded above by $1/p_1(x, z)$; hence, $Q^{x,z}$ is also absolutely continuous with respect to P^x on each $\mathcal{G}_n$.

Exercise 9.7.2 Show that the Radon-Nikodym derivative of the restriction of $Q^{x,z}$ to the σ-algebra $\mathcal{G}_n$ is

$$L_n = \left(\frac{dQ^{x,z}}{dP^x}\right)_{\mathcal{G}_n} = \frac{p_{n-1}(z, X_n)}{p_n(x, X_n)}. \tag{9.7.1}$$

By Example 8.5.2, the sequence $(L_n)_{n\geq 0}$ of Radon-Nikodym derivatives is a reverse martingale under the measure P^x. Therefore, the reverse Martingale Convergence Theorem (cf. Theorem 8.5.6 and Corollary 8.5.7) implies that with P^x-probability one,

$$\lim_{n\to\infty} L_n = L_\infty = \left(\frac{dQ^{x,z}}{dP^x}\right)_{\mathcal{T}} \quad P^x - \text{almost surely and in } L^1. \tag{9.7.2}$$

By hypothesis, the tail σ-algebra is trivial under P^x, so every $\mathcal{T}$-measurable random variable is P^x-almost surely constant; consequently, $P^x\{L_\infty = \alpha_x\} = 1$ for some constant $\alpha_x \in \mathbb{R}$. In fact, this constant must be $\alpha_x = 1$, because L_∞ is the likelihood ratio of a probability measure, and so $E^x L_\infty = 1$. Therefore, by (9.7.2), for any $x \in \Theta$,

$$\lim_{n\to\infty} E^x |L_n - L_\infty| = \lim_{n\to\infty} E^x |L_n - 1| = 0. \tag{9.7.3}$$

Relation (9.7.3), together with the formula (9.7.1), implies that the total variation distance between the probability distributions $p_{n-1}(z,\cdot)$ and $p_n(x,\cdot)$ converges to 0 as $n \to \infty$. In particular, by (9.7.3),

$$\sum_{y\in\Theta} |p_{n-1}(z,y) - p_n(x,y)| = \sum_{y\in\Theta} \left| \frac{p_{n-1}(z,y)}{p_n(x,y)} - 1 \right| p_n(x,y) = E^x|L_n - 1| \longrightarrow 0.$$

Laziness implies that the total variation distance between $p_{n-1}(z,\cdot)$ and $p_n(z,\cdot)$ converges to 0 (cf. relation (9.4.6)), so it follows by the triangle inequality that

$$\lim_{n\to\infty} \sum_{y\in\Theta} |p_n(z,y) - p_n(x,y)| = 0. \tag{9.7.4}$$

Finally, since the Markov chain is irreducible, any two states can be connected by a finite sequence of states along which the 1-step transition probabilities are positive; therefore, by another use of the triangle inequality, relation (9.7.4) holds for every pair of states x, z. □

Additional Notes. The cornerstone of the theory of bounded harmonic functions for Markov chains is Theorem 9.1.5, which is due to D. Blackwell [13]. The representation of bounded harmonic functions for the special case of random walks on free groups (Exercise 9.3.2) was first established by Dynkin and Maliutov [36]. For more on Pólya's Urn (Example 9.3.4), see Blackwell and Kendall [15].

The use of *coupling* in the study of Markov chains seems to have originated in the work of W. Doeblin [33]. The hard implication of Theorem 9.5.5 is essentially due to Blackwell and Freedman [14].

Chapter 10
Entropy

10.1 Avez Entropy and the Liouville Property

In Chapter 9 we showed that the Liouville property is equivalent to triviality of the invariant σ-algebra, and that this in turn is equivalent to weakly ergodicity of the underlying Markov chain. For random walks on finitely generated groups there is another equivalent condition.

Theorem 10.1.1 *A random walk on a finitely generated group whose Avez entropy is finite has the Liouville property if and only if its Avez entropy is* 0*.*

This theorem is an amalgamation of results due to Avez [4], Derriennic [32], and Kaimanovich & Vershik [68]. The bulk of this chapter will be devoted to its proof. Together with results we have already proved in chapters 3, 4, and 5, Theorem 10.1.1 leads to a number of interesting conclusions.

Corollary 10.1.2 *Irreducible, symmetric, random walks on nonamenable groups with finite Avez entropy always have nonconstant bounded harmonic functions, and therefore are never weakly ergodic.*

This follows directly from Kesten's theorem, which implies that any irreducible, symmetric random walk on a nonamenable group has positive entropy, and Theorem 9.5.5.

Corollary 10.1.3 *A nearest-neighbor random walk on a group of subexponential growth (i.e., a group for which $\beta = 0$, where β is defined by equation (3.2.1)) has no nonconstant bounded harmonic functions, and consequently is weakly ergodic.*

For nearest-neighbor random walks, this subsumes the results obtained by coupling arguments in Proposition 9.6.7 and Exercise 9.6.9; however, those results, which were proved using coupling arguments, did not require that the step distribution have finite support or even finite Shannon entropy.

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_10

10.2 Shannon Entropy and Conditional Entropy

The remainder of this chapter will be devoted to the proof of Theorem 10.1.1. The first order of business is a short course in the theory of Shannon entropy.

Definition 10.2.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\alpha = \{F_i\}_{1 \leq i \leq I}$ and $\beta = \{G_j\}_{1 \leq j \leq J}$ be two finite or countable measurable partitions of Ω (that is, partitions of Ω whose elements are contained in the sigma algebra $\mathcal{F}$). Define the *entropy* (or *Shannon entropy*) $H(\alpha)$ of the partition α and the *conditional entropy* $H(\alpha \mid \beta)$ of α given β by

$$H(\alpha) := -\sum_{i=1}^{I} P(F_i) \log P(F_i) \quad \text{and} \tag{10.2.1}$$

$$\begin{aligned} H(\alpha \mid \beta) &:= -\sum_{i=1}^{I} \sum_{j=1}^{J} P(F_i \cap G_j) \log P(F_i \mid G_j) \\ &= -\sum_{i=1}^{I} \sum_{j=1}^{J} P(F_i \cap G_j (\log P(F_i \cap G_j) - \log P(G_j)) \end{aligned} \tag{10.2.2}$$

Note: By convention, $0 \log 0 = 0$; consequently, sets of probability zero (in either of the partitions) have no effect on entropy or conditional entropy. We use the notation $P(F \mid G)$ for the usual naive conditional probability of F given G, that is, $P(F \mid G) = P(F \cap G)/P(G)$ unless $P(G) = 0$. The second expression for $H(\alpha)$ in (10.2.2) gives an equivalent definition:

$$H(\alpha \mid \beta) = H(\alpha \vee \beta) - H(\beta), \tag{10.2.3}$$

where $\alpha \vee \beta$ is the *join* of the partitions α and β, that is, the partition consisting of the non-empty sets in the list $\{F_i \cap G_j\}_{i \leq I, j \leq J}$. The alternative definition (10.2.3) is the natural analogue for entropy of the multiplication law $P(A \mid B)P(B) = P(A \cap B)$ for conditional probability. The notation $H(\mu)$ is also used for the Shannon entropy of a probability measure μ on a finite or countable set (cf. Corollary 3.3.3); in this chapter, $H(\cdot)$ will always refer to the entropy of a *partition.*

Proposition 10.2.2 *Shannon entropy is a* monotone *function of measurable partitions with respect to the refinement partial ordering, that is, if α is a refinement of β then $H(\alpha) \geq H(\beta)$, with equality if and only if the partitions α and β are identical modulo sets of probability* 0*. Consequently, $H(\alpha \mid \beta) = 0$ if and only if β is a refinement of α (up to changes by events of probability zero).*

Terminology: A partition α is said to be a *refinement* of β if every element of β is a union of elements of α.

Proof. If α is a refinement of β then $\alpha \vee \beta = \alpha$, and so $H(\alpha) - H(\beta) = H(\alpha \mid \beta) \geq 0$. The definition (10.2.2) shows that $H(\alpha \mid \beta) = 0$ if and only if $P(F_i | G_j) = 1$ for every pair $F_i \in \alpha$ and $G_j \in \beta$ such that $P(F_i \cap G_j) > 0$. Since α is by hypothesis a refinement of β, it follows that $P(F_i \Delta G_j) = 0$, so F_i and G_j differ by at most a set of measure 0. □

Proposition 10.2.3 *Let α, β, γ be measurable partitions such that γ is a refinement of β. Then*

$$H(\alpha) \geq H(\alpha \mid \beta) \geq H(\alpha \mid \gamma). \tag{10.2.4}$$

Furthermore, the addition law

$$H(\alpha) = H(\alpha \mid \beta) \quad \Longleftrightarrow \quad H(\alpha \vee \beta) = H(\alpha) + H(\beta) \tag{10.2.5}$$

holds if and only if the partitions α and β are independent.

Proof. Both inequalities (10.2.4) are consequences of Jensen's inequality (cf. Royden [111], Chapter 5) and the fact that the function $x \mapsto -x \log x$ is strictly concave on the unit interval. In particular, Jensen implies that for each element $F_i \in \alpha$,

$$\begin{aligned}
-\sum_j & P(F_i \cap G_j) \log P(F_i \mid G_j) \\
&= -\sum_j P(G_j) P(F_i \mid G_j) \log P(F_i \mid G_j) \\
&\leq -\left(\sum_j P(G_j) P(F_i \mid G_j)\right) \log \left(\sum_j P(G_j) P(F_i \mid G_j)\right) \\
&= -P(F_i) \log P(F_i).
\end{aligned}$$

Summing over i gives the inequality $H(\alpha) \geq H(\alpha \mid \beta)$. The second inequality is similar.

Jensen's inequality $E\varphi(X) \leq \varphi(EX)$ for a strictly concave function φ is strict unless the random variable X is constant. In the application above, the random variable is the function $j \mapsto P(F_i \mid G_j)$, with each j given probability $P(G_j)$. Thus, strict inequality holds unless for each i the conditional probabilities $P(F_i \mid G_j)$, where j ranges over the partition β, are all the same. But this will be the case only when $P(F_i \mid G_j) = P(F_i)$ for every pair i, j, that is, if the partitions are independent. □

Notational Convention: For any discrete random variable Y taking values in a finite or countable set $\{y_i\}_{i \leq I}$, let $\pi(Y)$ be the measurable partition $\{\{Y = y_i\}\}_{i \leq I}$. For any finite collection of discrete random variables $Y_1, Y_2, \cdots, Y_n$, all taking values

in finite sets, let

$$\pi((Y_i)_{1\leq i\leq n}) = \pi(Y_1, Y_2, \cdots, Y_n) = \bigvee_{i=1}^{n} \pi(Y_i)$$

be the join of the partitions $\pi(Y_j)$.

10.3 Avez Entropy

Now let's return to the world of random walks. Assume that under the probability measure P^x the sequence $X_n = X_0\xi_1\xi_2\cdots\xi_n$ is an irreducible random walk on a finitely generated group Γ with step distribution μ and initial state $X_0 = x$.

Assumption 10.3.1 *In this section, all entropies and conditional entropies are computed under $P = P^1$. We assume throughout that the step distribution of the random walk has finite Shannon entropy.*

This assumption ensures, by Proposition 10.2.3, that for every $n \in \mathbb{N}$,

$$H(\pi(\xi_1, \xi_2, \cdots, \xi_n)) = nH(\pi(\xi_1))).$$

Hence, by the monotonicity and addition laws for conditional entropy (Propositions 10.2.2 and 10.2.3), it follows that the conditional entropies in the propositions below are well-defined and finite.

Lemma 10.3.2 *For any integers $k, m \geq 1$ and $n \geq k$,*

$$H(\pi(X_k) \,|\, \pi(X_n, X_{n+1}, \cdots, X_{n+m})) = H(\pi(X_k) \,|\, \pi(X_n)). \tag{10.3.1}$$

Proof. The key is that the partition $\pi(X_n, X_{n+1}, \cdots, X_{n+m})$ is identical to the partition $\pi(X_n) \vee \pi(\xi_{n+1}, \xi_{n+2}, \cdots, \xi_{n+m})$; this is useful because if $k \leq n$ then the partitions

$$\pi(X_k, X_n) \quad \text{and} \quad \pi(\xi_{n+1}, \xi_{n+2}, \cdots, \xi_{n+m})$$

are independent. Thus, by the "addition law" (10.2.5) for independent partitions,

$$\begin{aligned}
&H\left(\pi(X_k) \,|\, \pi((X_i)_{n\leq i\leq n+m})\right) \\
&\quad = H\left(\pi(X_k) \vee \pi((X_i)_{n\leq i\leq n+m})\right) - H\left(\pi((X_i)_{n\leq i\leq n+m})\right) \\
&\quad = H(\pi(X_k) \vee \pi(X_n) \vee \pi((\xi_i)_{n<i\leq n+m})) - H(\pi(X_n) \vee \pi((\xi_i)_{n<i\leq n+m})) \\
&\quad = (H(\pi(X_k) \vee \pi(X_n)) + H(\pi(\xi_{n+1}, \xi_{n+2}, \cdots, \xi_{n+m}))) \\
&\qquad - (H(\pi(X_n)) + H(\pi(\xi_{n+1}, \xi_{n+2}, \cdots, \xi_{n+m})))
\end{aligned}$$

$$= H(\pi(X_k) \vee \pi(X_n)) - H(\pi(X_n)$$
$$= H(\pi(X_k) \,|\, \pi(X_n)).$$

□

Corollary 10.3.3 $H(\pi(X_k) \,|\, \pi(X_n)) \leq H(\pi(X_k) \,|\, \pi(X_{n+1}))$.

Proof. This is a direct consequence of the refinement inequality (10.2.4) and equation (10.3.1), which imply that

$$\begin{aligned} H(\pi(X_k) \,|\, \pi(X_n)) &= H(\pi(X_k) \,|\, \pi(X_n) \vee \pi(X_{n+1})) \\ &\leq H(\pi(X_k) \,|\, \pi(X_{n+1})). \end{aligned}$$

□

Corollary 10.3.4 *For any integer $k \geq 1$, the Avez entropy h satisfies*

$$kh = H(\pi(X_k)) - \lim_{n\to\infty} H(\pi(X_k) \,|\, \pi(X_n)). \tag{10.3.2}$$

Proof. The previous corollary implies that the sequence $H(\pi(X_k) \,|\, \pi(X_n))$ is nondecreasing, so the limit in (10.3.2) exists. Now rewrite the conditional entropy using equation (10.2.3):

$$H(\pi(X_k) \,|\, \pi(X_n)) = H(\pi(X_k) \vee \pi(X_n)) - H(\pi(X_n)).$$

The partition $\pi(X_k) \vee \pi(X_n)$ coincides with $\pi(X_k) \vee \pi(\xi_{k+1}\xi_{k+2}\cdots\xi_n)$, so by the addition law for independent partitions,

$$\begin{aligned} H(\pi(X_k) \vee \pi(X_n)) &= H(\pi(X_k)) + H(\pi(\xi_{k+1}\xi_{k+2}\cdots\xi_n)) \\ &= H(\pi(X_k)) + H(\pi(X_{n-k})), \end{aligned}$$

the latter equality because the distribution of the product $\xi_{k+1}\xi_{k+2}\cdots\xi_n$ is the same as that of $\xi_1\xi_2\cdots\xi_{n-k}$. Consequently, for any $n \geq 2$

$$H(\pi(X_k) \,|\, \pi(X_{nk})) = H(\pi(X_k)) + H(\pi(X_{nk-k})) - H(\pi(X_{nk})),$$

which, in view of Corollary 10.3.3, implies that the sequence $H(\pi(X_{nk})) - H(\pi(X_{nk-k}))$ is nonincreasing in n. Now for any nonincreasing sequence $(a_n)_{n\geq 1}$ of real numbers,

$$\inf_{n\geq 1} a_n = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} a_m;$$

applying this to the sequence $a_n = H(\pi(X_{nk})) - H(\pi(X_{nk-k}))$, we find that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}(H(\pi(X_{mk}))-H(\pi(X_{mk-k})))=\lim_{n\to\infty}\frac{1}{n}H(\pi(X_{nk}))=kh,$$

and so the relation (10.3.2) follows. □

Exercise 10.3.5 Show that

$$kh = H\left(\pi((X_i)_{1\le i\le k})\right) - \lim_{n\to\infty} H\left(\pi((X_i)_{1\le i\le k}\,|\,\pi(X_n)\right).$$

10.4 Conditional Entropy on a σ-Algebra

In this section, we show how to extend the definition (10.2.2) of conditional entropy in a natural way to σ-algebras. This will lead to a formula for the Avez entropy of a random walk that directly involves the tail σ-algebra.

Definition 10.4.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\alpha = \{F_i\}_{1\le i\le I}$ a measurable partition, and $\mathcal{G}$ a σ-algebra contained in $\mathcal{F}$. Define the *conditional entropy* of α with respect to $\mathcal{G}$ by

$$H(\alpha\,|\,\mathcal{G}) := \inf_{\beta\subset\mathcal{G}} H(\alpha\,|\,\beta), \tag{10.4.1}$$

where the infimum is over all finite measurable partitions whose events are elements of the σ-algebra $\mathcal{G}$.

Remark 10.4.2 If the σ-algebra $\mathcal{G}$ is finite, then it is generated by a finite measurable partition γ, consisting of those nonempty elements of $\mathcal{G}$ that have no nontrivial subsets in $\mathcal{G}$. In this case the monotonicity law (10.2.4) implies that

$$H(\alpha\,|\,\mathcal{G}) = H(\alpha\,|\,\gamma).$$

Thus, (10.4.1) is a natural extension of the definition (10.2.2). For any σ-algebra $\mathcal{G}$, there must exist finite partitions $\beta_n \subset \mathcal{G}$ such that

$$H(\alpha\,|\,\mathcal{G}) =\downarrow \lim_{n\to\infty} H(\alpha\,|\,\beta_n); \tag{10.4.2}$$

by the monotonicity property (10.2.4), the partitions β_n can be chosen so that each is a refinement of its predecessor.

Properties of Conditional Entropy:

(H1) Monotonicity: If $\mathcal{G}_1 \subset \mathcal{G}_2$ then $H(\alpha\,|\,\mathcal{G}_1) \ge H(\alpha\,|\,\mathcal{G}_2)$.
(H2) Independence Law: $H(\alpha\,|\,\mathcal{G}) = H(\alpha)$ if and only if α is independent of $\mathcal{G}$.

(H3) Containment Law: $H(\alpha \mid \mathcal{G}) = 0$ if and only if every event in α differs from an event in $\mathcal{G}$ by an event of probability 0.

Properties (H1) and (H2) follow trivially from the definition and Proposition 10.2.3. The reverse implication in (H3) follows from Proposition 10.2.2, because if α is almost surely equal to a partition $\alpha' \subset \mathcal{G}$ then $H(\alpha \mid \mathcal{G}) = H(\alpha' \mid \mathcal{G}) \leq H(\alpha' \mid \alpha') = 0$. The forward implication will follow from the next result.

Proposition 10.4.3 *For any measurable partition α and any σ-algebra $\mathcal{G} \subset \mathcal{F}$,*

$$H(\alpha \mid \mathcal{G}) = -\sum_{i=1}^{I} E\mathbf{1}_{F_i} \log(E(\mathbf{1}_{F_i} \mid \mathcal{G})) = -\sum_{i=1}^{I} EE(\mathbf{1}_{F_i} \mid \mathcal{G})(\log(E(\mathbf{1}_{F_i} \mid \mathcal{G})) \tag{10.4.3}$$

where $\alpha = \{F_i\}_{1 \leq i \leq I}$ and $E(\cdot \mid \mathcal{G})$ denotes the conditional expectation operator for the σ-algebra $\mathcal{G}$ and the probability measure P (cf. Section A.9 of the Appendix).

Proof. The second equality follows from the definition of conditional expectation. To prove the initial identity, consider first the case where $\mathcal{G}$ is finite. In this case $\mathcal{G}$ is generated by a finite partition $\gamma \subset \mathcal{G}$, and so $H(\alpha \mid \mathcal{G}) = H(\alpha \mid \gamma)$. Let $\gamma = \left\{G_j\right\}_{1 \leq j \leq J}$; then by an elementary calculation (see Exercise A.9.2 of the Appendix), for each set $F_i \in \alpha$,

$$E(\mathbf{1}_{F_i} \mid \mathcal{G}) = \sum_{j=1}^{J} \frac{P(F_i \cap G_j)}{P(G_j)} \mathbf{1}_{G_j}, \quad \text{and so}$$

$$-\sum_{i=1}^{I} E\mathbf{1}_{F_i} \log(E(\mathbf{1}_{F_i} \mid \mathcal{G})) = -E\sum_{i=1}^{I} \mathbf{1}_{F_i}\mathbf{1}_{G_j} \log \frac{P(F_i \cap G_j)}{P(G_j)} = H(\alpha \mid \gamma).$$

Thus, when $\mathcal{G}$ is finite the identity (10.4.3) is simply a reformulation of definition (10.2.2).

Now consider the general case, where the σ-algebra $\mathcal{G}$ might be infinite. The definition (10.4.1) implies that there is a sequence β_n of finite partitions contained in $\mathcal{G}$ such that $H(\alpha \mid \beta_n) \downarrow H(\alpha \mid \mathcal{G})$. Without loss of generality, we can arrange that each partition β_n is a refinement of the preceding partition β_{n-1}, because by relation (10.2.4),

$$H(\alpha \mid \mathcal{G}) = \lim_{n\to\infty} H(\alpha \mid \beta_n) \quad \Longrightarrow \quad H(\alpha \mid \mathcal{G}) = \lim_{n\to\infty} H\left(\alpha \mid \bigvee_{j=1}^{n} \beta_j\right).$$

Thus, the finite σ-algebras $\mathcal{G}_n := \sigma(\beta_n)$ form a nested sequence $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \cdots \subset \mathcal{G}$, and so for each set F_i in the partition α, the sequence $E(\mathbf{1}_{F_i} \mid \mathcal{G}_n)$ is a bounded, nonnegative martingale (by the "Tower Property" of conditional expectation — see Section A.9 of the Appendix). Therefore, the Martingale Convergence Theorem

implies that for each $F_i \in \alpha$,

$$\lim_{n\to\infty} E(\mathbf{1}_{F_i} \mid \mathcal{G}_n) = E(\mathbf{1}_{F_i} \mid \mathcal{G}_\infty)$$

almost surely and in L^1, where $\mathcal{G}_\infty = \sigma(\cup_{n\geq 1}\mathcal{G}_n) \subset \mathcal{G}$. Since the identity (10.4.3) holds for each of the σ-algebras $\mathcal{G}_n$, it follows by the monotonicity property (H1) that

$$H(\alpha \mid \mathcal{G}) = H(\alpha \mid \mathcal{G}_\infty) = -\sum_{i=1}^{I} E\mathbf{1}_{F_i} \log E(\mathbf{1}_{F_i} \mid \mathcal{G}_\infty). \tag{10.4.4}$$

Thus, to complete the proof it will suffice to show that $E(\mathbf{1}_{F_i} \mid \mathcal{G}) = E(\mathbf{1}_{F_i} \mid \mathcal{G}_\infty)$ for each $i \leq I$.

The identity (10.4.4) holds for every $\mathcal{F}$-measurable partition, so in particular it holds for every two-element partition $\alpha = \{F, F^c\}$ with $F \in \mathcal{G}$. For any such partition $H(\alpha \mid \mathcal{G}) = 0$, by the (reverse implication in the) Containment Law (H3); hence,

$$E(\log E(\mathbf{1}_F \mid \mathcal{G}_\infty)) = 0 \quad \text{for every} \quad F \in \mathcal{G}.$$

This implies that $E(\mathbf{1}_F \mid \mathcal{G}_\infty)$ is either 0 or 1 with probability 1, which is only possible if there is an event $G \in \mathcal{G}_\infty$ such that $\mathbf{1}_F = \mathbf{1}_G$ almost surely (cf. Exercise A.9.5 in the Appendix). Now the definition of conditional expectation (cf. Exercise A.9.3) implies that if two σ-algebras $\mathcal{G}' \subset \mathcal{G}''$ agree up to null sets (that is, for every $G'' \in \mathcal{G}''$ there is an event $G' \in \mathcal{G}'$ such that $P(G' \Delta G'') = 0$), then

$$E(Z \mid \mathcal{G}') = E(Z \mid \mathcal{G}'') \quad \text{for all} \quad Z.$$

Therefore, for every measurable partition α the conditional expectations in (10.4.4) coincide (almost surely) with the corresponding conditional expectations in (10.4.3). □

Proof of the Containment Law (H3) $\Longrightarrow$ Suppose that α is a measurable partition such that $H(\alpha \mid \mathcal{G}) = 0$. The identity (10.4.3) then implies that for every $F \in \alpha$,

$$\log E(\mathbf{1}_F \mid \mathcal{G}) = 0 \quad \text{almost surely.}$$

By the linearity and monotonicity properties of conditional expectation, $0 \leq E(\mathbf{1}_F \mid \mathcal{G}) \leq 1$ almost surely, so it must be the case that for some event $G \in \mathcal{G}$,

$$E(\mathbf{1}_F \mid \mathcal{G}) = \mathbf{1}_G \quad \text{almost surely.}$$

But this implies that $\mathbf{1}_F = \mathbf{1}_G$ almost surely (see Exercise A.9.5). □

Corollary 10.4.4 (Continuity of Conditional Entropy) *Let α be any $\mathcal{F}$-measurable partition and let $(\mathcal{G}_n)_{n\in\mathbb{N}}$ be a sequence of σ-algebras all contained in $\mathcal{F}$.*

(H4) If $\mathcal{G}_1 \supset \mathcal{G}_2 \supset \mathcal{G}_3 \supset \cdots$ then $\uparrow \lim_{n\to\infty} H(\alpha \,|\, \mathcal{G}_n) = H(\alpha \,|\, \cap_{n\in\mathbb{N}} \mathcal{G}_n)$.
(H5) If $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{G}_3 \subset \cdots$ then $\downarrow \lim_{n\to\infty} H(\alpha \,|\, \mathcal{G}_n) = H(\alpha \,|\, \sigma(\cup_{n\in\mathbb{N}}\mathcal{G}_n))$.

Proof. For any event F the sequence of conditional expectations $E(\mathbf{1}_F \,|\, \mathcal{G}_n)$ is a bounded martingale (if the σ-algebras $\mathcal{G}_n$ increase) or reverse martingale (if they decrease), and so the martingale convegence theorems imply that they converge a.s. and (by bounded convergence) in L^1. Therefore, the results (H4)-(H5) follow from the formula (10.4.3) for conditional entropy. $\square$

10.5 Avez Entropy and Boundary Triviality

Armed with the notion of entropy conditional on a σ-algebra, we can now directly relate the Avez entropy $h = h(\mu; \Gamma)$ to conditional entropies given the tail σ-algebra.

Corollary 10.5.1 (Avez Entropy) *Let $(\Omega, \mathcal{F}, P)$ be a probability space supporting a lazy random walk $\mathbf{X} = (X_n)_{n\geq 0}$ on a finitely generated group Γ whose step distribution μ has finite Shannon entropy, and denote by $\pi(X_i)$ the measurable partition generated by X_i. The Avez entropy $h = h(\Gamma; \mu)$ of the random walk satisfies*

$$\begin{aligned} h &= H(\pi(X_1)) - H(\pi(X_1) \,|\, \mathbf{X}^{-1}(\mathcal{T})) \\ &= H(\pi(X_1)) - H(\pi(X_1) \,|\, \mathbf{X}^{-1}(\mathcal{I})) \end{aligned} \tag{10.5.1}$$

where $\mathcal{T}$ and $\mathcal{I}$ are the tail and invariant σ-algebras on Γ^∞, respectively (cf. Sections 9.1 and 9.4). More generally, for any integer $k \geq 1$,

$$\begin{aligned} kh &= H(\pi(X_k)) - H(\pi(X_k) \,|\, \mathbf{X}^{-1}(\mathcal{T})) \\ &= H(\pi(X_k)) - H(\pi(X_k) \,|\, \mathbf{X}^{-1}(\mathcal{I})) \end{aligned} \tag{10.5.2}$$

Proof. For a lazy random walk, the tail σ-algebra $\mathbf{X}^{-1}(\mathcal{T})$ is a null extension of the invariant σ-algebra $\mathbf{X}^{-1}(\mathcal{I})$, by Proposition 9.4.3. Consequently, by Exercise A.9.3, for every event F the conditional expectations $E(\mathbf{1}_F \,|\, \mathbf{X}^{-1}(\mathcal{T}))$ and $E(\mathbf{1}_F \,|\, \mathbf{X}^{-1}(\mathcal{I}))$ are equal almost surely, and so by Proposition 10.4.3,

$$H(\pi(X_k) \,|\, \mathbf{X}^{-1}(\mathcal{T})) = H(\pi(X_k) \,|\, \mathbf{X}^{-1}(\mathcal{I})).$$

Thus, it suffices to prove the first equality in (10.5.2).

By Lemma 10.3.2, for any integers $k, m \geq 1$ and $n \geq k$,

$$\begin{aligned} H(\pi(X_k) \mid \pi(X_n)) &= H\left(\pi(X_k) \mid \bigvee_{i=n}^{n+m} \pi(X_i)\right) \\ &= H(\pi(X_k) \mid \sigma(X_i)_{n \le i \le n+m}) \end{aligned}$$

and so by the Continuity Property (H5) of Corollary 10.4.4,

$$H(\pi(X_k) \mid \pi(X_n)) = H(\pi(X_k) \mid \sigma(X_i)_{i \ge n}).$$

Therefore, by Corollary 10.3.4,

$$\begin{aligned} kh &= H(\pi(X_k)) - \lim_{n\to\infty} H(\pi(X_k) \mid \pi(X_n)) \\ &= H(\pi(X_k)) - \lim_{n\to\infty} H(\pi(X_k) \mid \sigma(X_i)_{i \ge n}) \\ &= H(\pi(X_k)) - H(\pi(X_k) \mid \mathbf{X}^{-1}(\mathcal{T})), \end{aligned}$$

the last by the Continuity Property (H4). □

Proof of Theorem 10.1.1 It suffices to prove the theorem for *lazy* random walks. To see this, recall that for any random walk, any lazy variant has the same bounded harmonic functions, so one has the Liouville property if and only if the other does. Furthermore, a random walk has positive speed if and only if every lazy variant has positive speed (cf. Exercise 3.2.8); consequently, by Corollary 4.1.2 and Proposition 4.1.3, a random walk has positive Avez entropy if and only if every lazy variant also has positive Avez entropy.

Suppose that the random walk has the Liouville property. Then by Theorem 9.1.5, the invariant σ-algebra $\mathbf{X}^{-1}(\mathcal{I})$ is trivial, in the sense that every invariant event has probability either 0 or 1. But if this is the case, the σ-algebra $\mathbf{X}^{-1}(\mathcal{I})$ is independent of every measurable partition, so by the Independence Law (H2) for conditional entropy,

$$H(\pi(X_1) \mid \mathbf{X}^{-1}(\mathcal{I})) = H(\pi(X_1)).$$

By Corollary 10.5.1, this implies that $h = 0$.

Now suppose conversely that $h = 0$. Then by Corollary 10.5.1, for every $k \ge 1$

$$0 = H(\pi(X_k)) - H(\pi(X_k) \mid \mathbf{X}^{-1}(\mathcal{T})).$$

Thus, by the Independence Law (H2) for conditional entropy, the tail σ-algebra $\mathbf{X}^{-1}(\mathcal{T})$ is independent of the partition $\pi(X_k)$, and hence also independent of the entire σ-algebra $\sigma(X_k)$, for every $k \ge 1$. The events in these σ-algebras generate the σ-algebra $\mathcal{F}_\infty := \sigma(X_1, X_2, \cdots)$, so it follows that every event in $\mathcal{F}_\infty$ is independent of the tail σ-algebra $\mathbf{X}^{-1}(\mathcal{T})$. Hence, since σ-algebra is contained in $\mathcal{F}_\infty$, the tail σ-algebra is independent of itself, in particular, for every event $G \in \mathbf{X}^{-1}(\mathcal{T})$,

$$P(G) = P(G \cap G) = P(G)^2.$$

Therefore, the σ-algebra $\mathbf{X}^{-1}(\mathcal{T})$ is trivial, in the sense that every event in $\mathbf{X}^{-1}(\mathcal{T})$ has probability either 0 or 1. Since the invariant σ-algebra $\mathbf{X}^{-1}(\mathcal{I})$ is contained in the tail σ-algebra $\mathbf{X}^{-1}(\mathcal{T})$, it is also trivial, and so by Theorem 9.1.5, the random walk has the Liouville property. $\square$

10.6 Entropy and Kullback-Leibler Divergence

There is an equivalent formulation of Avez entropy that emphasizes the role of the exit measures ν_z defined in Section 9.1. This involves the notion of *Kullback-Leibler divergence*.

Definition 10.6.1 If λ_1, λ_2 are probability measures on $(\Omega, \mathcal{G})$ such that λ_1 is absolutely continuous with respect to λ_2, then the *Kullback-Leibler divergence* $D(\lambda_2 \parallel \lambda_1)$ is defined to be the nonnegative real number

$$D(\lambda_2 \parallel \lambda_1) := -\int \log \left(\frac{d\lambda_1}{d\lambda_2}\right)_{\mathcal{G}} d\lambda_2 \tag{10.6.1}$$

where $(d\lambda_1/d\lambda_2)_{\mathcal{G}}$ is the likelihood ratio of λ_1 with respect to λ_2 on the σ-algebra $\mathcal{G}$ (cf. Section A.8 of the Appendix).

Lemma 10.6.2 *The Kullback-Leibler divergence $D(\lambda_2 \parallel \lambda_1)$ is nonnegative (but possibly infinite), and strictly positive unless $\lambda_1 = \lambda_2$.*

Proof. The function $-\log x$ is strictly convex on $\mathbb{R}_+$, so Jensen's inequality implies that

$$-\int \log \left(\frac{d\lambda_1}{d\lambda_2}\right)_{\mathcal{G}} d\lambda_2 \geq -\log \int \left(\frac{d\lambda_1}{d\lambda_2}\right)_{\mathcal{G}} d\lambda_2 = 0.$$

The inequality is strict unless the integrand is λ_2-almost surely constant, which will be the case if and only if $\lambda_1 = \lambda_2$. $\square$

Exercise 10.6.3 Let $T : \Omega \to \Omega$ be an invertible, $\mathcal{G}$-measurable transformation. Show that if $\lambda_1 \ll \lambda_2$ then $\lambda_1 \circ T^{-1} \ll \lambda_2 \circ T^{-1}$ and

$$D(\lambda_2 \parallel \lambda_1) = D(\lambda_2 \circ T^{-1} \parallel \lambda_1 \circ T^{-1}). \tag{10.6.2}$$

More generally, let $T : \Omega \to \Omega'$ be a mapping from one measurable space $(\Omega, \mathcal{G})$ to another $(\Omega', \mathcal{G}')$ such that $\mathcal{G} = T^{-1}(\mathcal{G}')$. Show that if λ_1, λ_2 are probability measures on $\mathcal{G}$ such that $\lambda_1 \ll \lambda_2$, then $\lambda_1 \circ T^{-1} \ll \lambda_2 \circ T^{-1}$ and the identity (10.6.2) holds.

HINT: Exercises A.8.7 and A.8.8 of the Appendix might be helpful.

Exercise 10.6.4 Let λ_1, λ_2 be probability measures on $(\Omega, \mathcal{G})$ such that λ_1 is absolutely continuous with respect to λ_2, and let $\mathcal{H} \subset \mathcal{G}$ be a σ-algebra contained in $\mathcal{G}$. Denote by $\lambda_i \restriction \mathcal{H}$ the restriction of the measure λ_i to the σ-algebra $\mathcal{H}$. Prove that

$$D(\lambda_2 \restriction \mathcal{H} \,\|\, \lambda_1 \restriction \mathcal{H}) \leq D(\lambda_2 \restriction \mathcal{G} \,\|\, \lambda_1 \restriction \mathcal{G}), \tag{10.6.3}$$

with strict inequality unless the likelihood ratio $\left(\frac{d\lambda_1}{d\lambda_2}\right)_{\mathcal{G}}$ is (up to change on an event of ν_1-probability zero) $\mathcal{H}$-measurable.

HINT: The Radon-Nikodym Theorem and the definition of conditional expectation imply that

$$\left(\frac{d\nu_1}{d\nu_2}\right)_{\mathcal{H}} = E_{\nu_2}\left(\left(\frac{d\nu_1}{d\nu_2}\right)_{\mathcal{G}} \,\middle|\, \mathcal{H}\right).$$

Use this in conjunction with Jensen's inequality for conditional expectation (cf. Theorem A.9.7 of the Appendix.)

Definition 10.6.5 For any event $F \in \mathcal{F}$ of positive probability in a probability space $(\Omega, \mathcal{F}, P)$ and any σ-algebra $\mathcal{G} \subset \mathcal{F}$, define the *conditional measure* Q_F on $\mathcal{G}$ by

$$Q_F(G) = \frac{P(F \cap G)}{P(F)} = \frac{E_P \mathbf{1}_F \mathbf{1}_G}{P(F)}. \tag{10.6.4}$$

Proposition 10.6.6 *For any $\mathcal{F}$-measurable partition α and any σ-algebra $\mathcal{G} \subset \mathcal{F}$,*

$$H(\alpha \,|\, \mathcal{G}) = H(\alpha) - \sum_{F \in \alpha} P(F) D(Q_F \,\|\, P). \tag{10.6.5}$$

Note: The measure Q_F is undefined for those F such that $P(F) = 0$, so we adopt the convention that for such events $P(F)D(Q_F \,\|\, P) = 0$.

Proof. For any event $F \in \mathcal{F}$ of positive probability the conditional measure Q_F is absolutely continuous with respect to P, and by definition of conditional expectation, the likelihood ratio of Q_F with respect to P on $\mathcal{G}$ is

$$\left(\frac{dQ_F}{dP}\right)_{\mathcal{G}} = E_P(\mathbf{1}_F \,|\, \mathcal{G})/P(F). \tag{10.6.6}$$

Thus, by Proposition 10.4.3,

$$
\begin{aligned}
H(\alpha \mid \mathcal{G}) &= -\sum_{F\in\alpha} E_P \mathbf{1}_F \log(E_P(\mathbf{1}_F \mid \mathcal{G})) \\
&= -\sum_{F\in\alpha} P(F) E_{Q_F} \log(E_P(\mathbf{1}_F \mid \mathcal{G})) \\
&= -\sum_{F\in\alpha} P(F) E_{Q_F} \log(dQ_F/dP)_{\mathcal{G}} - \sum_{F\in\alpha} P(F) \log P(F) \\
&= -\sum_{F\in\alpha} P(F) D(Q_F \parallel P) + H(\alpha).
\end{aligned}
$$

□

Corollary 10.6.7 *The Avez entropy $h = h(\Gamma; \mu)$ of an aperiodic random walk $\mathbf{X} = (X_n)_{n\geq 0}$ on a finitely generated group Γ with finitely supported step distribution μ satisfies*

$$
nh = \sum_{z\in\Gamma} \mu^{*n}(z) D(\nu_z \parallel \nu) = \sum_{z\in\Gamma} \mu^{*n}(z) D(\nu \parallel \nu_{z^{-1}}) \quad \forall n \in \mathbb{N} \tag{10.6.7}
$$

where ν_g and $\nu = \nu_1$ are the exit measures of the random walk with initial points $X_0 = g$ and $X_0 = 1$, respectively.

Proof. The identity $D(\nu_z \parallel \nu) = D(\nu \parallel \nu_{z^{-1}})$ follows from Exercise 10.6.3, so it suffices to prove the first equality in (10.6.7). For this, observe that for any event $F = \{X_n = z\}$ in the partition $\pi(X_n)$, the conditional measure Q_F on the invariant σ-algebra $\mathbf{X}^{-1}(\mathcal{I})$ is related to the exit measure ν_z by

$$
\nu_z = Q_F \circ \mathbf{X}^{-1}.
$$

This follows from the Markov property: for any event $G \in \mathcal{I}$, with $F = \{X_n = z\}$

$$
\begin{aligned}
Q_F \circ \mathbf{X}^{-1}(G) &= P(\{(X_0, X_1, \cdots) \in G\} \cap F)/P(F) \\
&= P(\{(X_n, X_{n+1}, \cdots) \in G\} \cap F)/P(F) \quad \text{(by invariance of } G) \\
&= P^z\{(X_0, X_1, \cdots) \in G\} \\
&= \nu_z(G).
\end{aligned}
$$

Similarly, $P \circ \mathbf{X}^{-1} = \nu$. Therefore, by Proposition 10.6.6 and Corollary 10.5.1,

$$
\begin{aligned}
nh &= H(\pi(X_n)) - H(\pi(X_n) \mid \mathbf{X}^{-1}(\mathcal{I})) \\
&= \sum_{z\in\Gamma} \mu^{*n}(z) D(Q_{\{X_n=z\}} \upharpoonright \mathbf{X}^{-1}(\mathcal{I}) \parallel P \upharpoonright \mathbf{X}^{-1}(\mathcal{I})) \\
&= \sum_{z\in\Gamma} \mu^{*n}(z) D(\nu_z \parallel \nu),
\end{aligned}
$$

the last by equation (10.6.2). □

Additional Notes. Avez [5] proved that zero entropy implies the Liouville property. The reverse implication was proved independently by Derriennic [32] and Kaimanovich & Vershik [68]. For various classes of groups of subexponential growth, the Liouville property was established long before Theorem 10.1.1. For $\Gamma = \mathbb{Z}^d$, it was first proved by Blackwell [13]. Choquet and Deny [24] later showed that Blackwell's theorem extends to random walk on any abelian group.

Following [43], say that a finitely generated group Γ has the *(Blackwell)-Choquet-Deny property* if *every* nondegenerate random walk on Γ has the Liouville property. Which finitely generated groups have the Blackwell-Choquet-Deny property? In a tour de force, Frisch, Hartman, Tamuz, and Ferdowski [43] have recently proved that these are precisely the *virtually nilpotent* groups (see Section 15.5 below for the relevant definitions).

Chapter 11
Compact Group Actions and Boundaries

Geometric group theory is devoted to the interplay between the algebraic structure of a group and properties of its actions on geometric and topological objects. We have seen that the action of a finitely generated group on its Cayley graph by right multiplication is of fundamental importance in the study of random walks. In this chapter we will investigate group actions on *compact* metric spaces, and show how these can unlock information about the long-time behavior of random walk trajectories.

11.1 Γ-Spaces

Recall (cf. Definition 1.2.4) that an *action* (or *topological action*) of a finitely generated group (or matrix group) Γ on a topological space $\mathcal{Y}$ is a homomorphism Φ from Γ to the group of homeomorphisms of $\mathcal{Y}$. For the homeomorphism $\Phi(g)$ associated with a group element g we usually use the shorthand notation $y \mapsto g \cdot y$.

Definition 11.1.1 A *compact* Γ*-space* is a compact metric space $(\mathcal{Y}, d)$ paired with a topological Γ-action (that is, an action by homeomorphisms).

Example 11.1.2 Let $\mathbb{F}_k$ be the free group on $k \geq 2$ generators $\{a_1, a_2, \cdots, a_k\}$, and let $\mathbb{A} = \left\{a_i^{\pm 1}\right\}_{i \leq k}$ be the natural symmetric generating set. Define $\partial \mathbb{F}_k$ to be the set of all infinite reduced words with entries in the alphabet $\mathbb{A}$. For any two elements $\omega, \omega' \in \partial \mathbb{F}_k$, define

$$d(\omega, \omega') = 2^{-n(\omega,\omega')},$$

where $n(\omega, \omega') \geq 0$ is the maximal integer n such that the sequences ω and ω' agree in their first n coordinates (or ∞ if $\omega- = \omega'$). Then $(\partial \mathbb{F}_k, d)$ is a compact metric space (cf. Exercise 1.6.4); the topology induced by the metric d is the topology

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_11

of coordinatewise convergence. The group $\mathbb{F}_k$ acts on $\partial\mathbb{F}_k$ by left concatenation followed by reduction, e.g., if $k = 2$ and $\mathbb{A} = \{a^{\pm 1}, b^{\pm 1}\}$ then

$$(aba) \cdot (aab^{-1}ab \cdots) = abaaab^{-1}ab \cdots \quad \text{and}$$
$$(aba) \cdot (a^{-1}bba^{-1}a^{-1} \cdots) = abbba^{-1}a^{-1} \cdots .$$

Henceforth, we will refer to this as the *natural action* of $\mathbb{F}_k$ on the boundary $\partial\mathbb{F}_k$.

Example 11.1.3 Let $\Gamma = SL(k, \mathbb{R})$ be the group of $k \times k$ real matrices with determinant 1 (or any subgroup), and let $S^{k-1} = \{v \in \mathbb{R}^k : \|v\| = 1\}$ be the unit sphere in $\mathbb{R}^k$. Invertible matrices map lines through the origin to lines through the origin, so matrix multiplication induces an action of $SL(k, \mathbb{R})$ on S^{k-1} by the following rule: for any $g \in SL(k, \mathbb{R})$ and any $v \in S^{k-1}$,

$$g \cdot v = \frac{gv}{\|gv\|}. \tag{11.1.1}$$

Since any invertible matrix g maps antipodal vectors $\pm v$ to antipodal vectors $\pm gv$, the action of matrix multiplication induces a corresponding multiplicative action on the quotient space $\mathbb{R}^k/\sim$, where $\sim$ denotes the equivalence relation $v \sim w$ if and only if $v = \pm w$, by

$$g(\pm v) = \pm(gv). \tag{11.1.2}$$

This in turn induces an action on the real projective space $\mathbb{RP}^{k-1} := S^{k-1}/\sim$ by

$$g \cdot (\pm v) = \pm\frac{gv}{\|gv\|}. \tag{11.1.3}$$

Remark 11.1.4 Why bother with the action (11.1.3) when it is "essentially the same" as the action (11.1.1)? The answer is that a matrix $g \in SL(2, \mathbb{R})$ with large norm $\|g\|$ nearly collapses the projective line $\mathbb{RP}^1$ to a single point: see Lemma 11.4.12 below for a precise statement of this property.

Example 11.1.5 † The group $SL(2, \mathbb{R})$ also acts by *linear fractional transformations* on the upper half plane $\mathbb{H} := \{z \in \mathbb{C} : \Im z > 0\}$, in detail,

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \cdot z := \frac{\alpha z + \beta}{\gamma z + \delta}. \tag{11.1.4}$$

See [1], Chapter 3, Section 3 for a brief introduction to the salient facts; in particular, each $g \in SL(2, \mathbb{R})$ acts *bijectively* on both $\mathbb{H}$ and its boundary $\partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$. The boundary $\partial\mathbb{H}$ is the one-point compactification of $\mathbb{R}$, and as such is homeomorphic to the unit circle S^1. Thus, the space $\partial\mathbb{H}$, when endowed with the $SL(2, \mathbb{R})$-action (11.1.4), is a compact Γ-space.

Example 11.1.6 † The group $GL(2,\mathbb{C})$ of invertible 2×2 complex matrices acts on the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C}\cup\{\infty\}$ by linear fractional transformations. Those elements of $GL(2,\mathbb{C})$ with determinant 1 and entries (11.1.4) satisfying $\gamma = \bar{\beta}$ and $\delta = \bar{\alpha}$ map the unit circle to itself, and therefore the unit disk $\mathbb{D}$ to itself; the subgroup of $GL(2,\mathbb{C})$ consisting of all such matrices is known as $SU(1,1)$. The action of $SU(1,1)$ on the unit disk is *conjugate* to the action (11.1.4) of $SL(2,\mathbb{R})$ on the upper half-plane, that is, there is an isomorphism $\Psi : SL(2,\mathbb{R}) \to SU(1,1)$ and a homeomorphism $F : \mathbb{H} \to \mathbb{D}$ such that for any $g \in SL(2,\mathbb{R})$ and $z \in \mathbb{H}$,

$$F(g\cdot z) = \Psi(g)\cdot(F(z)) \tag{11.1.5}$$

where $\Psi : SL(2,\mathbb{R}) \to SU(1,1)$ is the group isomorphism and $F : \mathbb{H} \to \mathbb{D}$ the conformal homeomorphism defined by

$$F(z) := \frac{z-i}{z+i} \quad \text{and} \quad \Phi(g) = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} g/(2i). \tag{11.1.6}$$

Fuchsian groups are discrete subgroups of $SU(1,1)$ (or equivalently, of $SL(2,\mathbb{R})$). The free group $\mathbb{F}_2$ on two generators is (isomorphic to) a Fuchsian group Γ; the generators are the linear transformations $A, B \in SU(1,1)$ such that

$$A\cdot(-1) = +i; \qquad A\cdot(-i) = +1,$$
$$B\cdot(-1) = -i; \qquad B\cdot(+i) = +1;$$

see Figure 11.1. The action of $\mathbb{F}_2$ on the unit disk is illustrated by the second pane of Figure 11.1. Images of the "rectangle" R bounded by the four quarter circles with endpoints $\pm 1, \pm i$ by the elements of Γ tessellate the disk as shown in Figure 11.1. Group elements are in one-to-one correspondence with "tiles" of the tessellation; two group elements are nearest neighbors in the Cayley graph (cf.

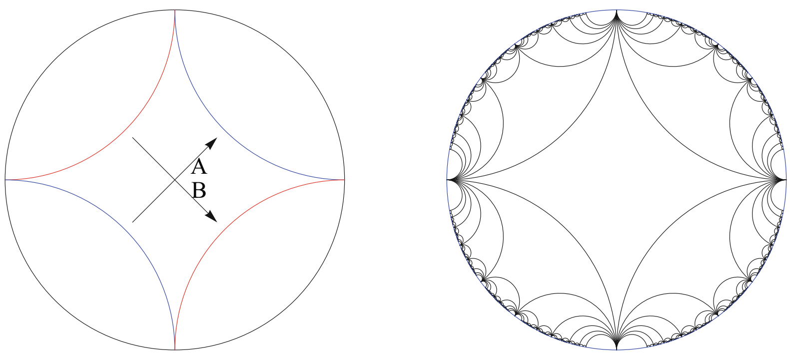


Fig. 11.1 Embedding $\mathbb{F}_2$ in $SU(1,1)$

figure in Example 1.2.7) if and only if the corresponding tiles share an edge in the tessellation.

11.2 Stationary and Invariant Measures

Definition 11.2.1 Let Φ be a topological action of a finitely generated group Γ on a compact metric space $\mathcal{Y}$. The *induced action* on the space $\mathcal{M}(\mathcal{Y})$ of Borel probability measures on $\mathcal{Y}$ is defined as follows: for any $\lambda \in \mathcal{M}(\mathcal{Y})$ and any $g \in \Gamma$,

$$g \cdot \lambda := \lambda \circ \Phi(g)^{-1} \tag{11.2.1}$$

is the pushforward of the measure λ by the (continuous, and hence Borel measurable) mapping $\Phi(g) : \mathcal{Y} \to \mathcal{Y}$. Equivalently, $g \cdot \lambda$ is the unique Borel probability measure such that for every continuous (and hence, every bounded, Borel measurable) function $f : \mathcal{Y} \to \mathbb{R}$,

$$\int f(y)\, d(g \cdot \lambda)(y) = \int f(g \cdot y) d\lambda(y). \tag{11.2.2}$$

For any probability measure μ on Γ and any $\lambda \in \mathcal{M}(\mathcal{Y})$, define the *convolution* $\mu * \lambda \in \mathcal{M}(\mathcal{Y})$ to be the convex combination

$$\mu * \lambda = \sum_{g \in \Gamma} \mu(g)(g \cdot \lambda). \tag{11.2.3}$$

A probability measure $\lambda \in \mathcal{M}(\mathcal{Y})$ is *μ-stationary* if $\lambda = \mu * \lambda$.

The equivalence of (11.2.1) and (11.2.2) follows by the Riesz-Markov Theorem: see Section A.7 of the Appendix for this and for the basics of the theory of weak convergence, which will be used freely in this chapter and the next.

The identity (11.2.1) can be reinterpreted in the language of probability theory as follows: if Y is a $\mathcal{Y}$-valued random variable with distribution λ, then $g \cdot \lambda$ is the distribution of the random variable $g \cdot Y$. Similarly, if X is a Γ-valued random variable with distribution μ such that X and Y are independent, then $\mu * \lambda$ is the distribution of $X \cdot Y$. Thus, if λ is μ-stationary and $Y, \xi_1, \xi_2, \cdots$ are independent random variables such that Y has distribution λ and $\xi_1, \xi_2, \cdots$ all have distribution μ, then the sequence

$$Y_n := \xi_n \xi_{n-1} \cdots \xi_1 \cdot Y$$

is stationary in the sense of Definition 2.1.5.

Exercise 11.2.2 Check that convolution is associative and distributive over convex combinations, that is, for any probability distributions μ, μ_1, μ_2 on Γ, any Borel

probability measures λ, λ' on $\mathcal{Y}$, and any scalar $\alpha \in [0, 1]$,

$$(\mu_1 * \mu_2) * \lambda = \mu_1 * (\mu_2 * \lambda) \quad \text{and}$$
$$\mu * (\alpha\lambda + (1-\alpha)\lambda') = \alpha\mu * \lambda + (1-\alpha)\mu * \lambda'.$$

Exercise 11.2.3 Let $\mathcal{Y}$ be a compact Γ-space, where Γ is a finitely generated group, and suppose that $(\lambda_m)_{m\in\mathbb{N}}$ are Borel probability measures on $\mathcal{Y}$ that converge weakly as $m \to \infty$ to a probability measure λ. Show that for any probability measure μ on Γ the convolutions $\mu * \lambda_m$ converge weakly to $\mu * \lambda$. Conclude that any weak limit of a sequence of μ-stationary probability measures is μ-stationary.

Proposition 11.2.4 *For any Γ-action on a compact metric space $\mathcal{Y}$ and any probability measure μ on Γ there exists a μ-stationary Borel probability measure λ on $\mathcal{Y}$.*

Proof. [1]This is a variation of a standard argument in dynamical systems due to Krylov and Bogolyubov [82]: the strategy is to construct a stationary measure by extracting a weakly convergent subsequence of a sequence of nearly-stationary Cesaro averages. Fix a point $y \in \mathcal{Y}$, and for each integer $n \geq 1$ define

$$\lambda_n = \frac{1}{n}\sum_{k=0}^{n-1} \mu^{*k} * \delta_y$$

where $\mu^{*k} * \delta_y$ is the convolution of δ_y, the unit point mass at y, with the kth convolution power of μ. Since $\mathcal{Y}$ is a compact and metrizable, the Helly Selection Principle (Theorem A.7.9 of the Appendix) guarantees that the sequence $(\lambda_n)_{n\geq 1}$ has a weakly convergent subsequence $(\lambda_{n_m})_{m\geq 1}$. Let $\lambda = \text{weak-}\lim \lambda_{n_m}$ be the limit measure. We will prove that λ is μ-stationary.

By Exercise 11.2.3, $\mu * \lambda$ is the weak limit of the (sub)sequence $\mu * \lambda_{n_m}$. Moreover, by Exercise 11.2.2, for each $n \in \mathbb{N}$,

$$\mu * \left(\frac{1}{n}\sum_{k=0}^{n-1} \mu^{*k} * \lambda\right) = \frac{1}{n}\sum_{k=1}^{n} \mu^{*k} * \lambda.$$

Therefore, for any continuous function $f : \mathcal{Y} \to \mathbb{R}$,

$$\int f\, d(\mu * \lambda) = \lim_{m\to\infty} \int f(y)\, d\left(\mu * \frac{1}{n_m}\sum_{k=0}^{n_m-1} \mu^{*k} * \lambda\right)(y)$$

[1] Proposition 11.2.4 also follows from the *Schauder-Tychonoff Fixed-Point Theorem*. See Zeidler [134], Chapter 2, for a discussion of this theorem and its uses.

$$
\begin{aligned}
&= \lim_{m\to\infty} \frac{1}{n_m} \sum_{k=1}^{n_m} \int f(y)d(\mu^{*k} * \lambda)(y) \\
&= \lim_{m\to\infty} \frac{1}{n_m} \sum_{k=0}^{n_m-1} \int f(y)d(\mu^{*k} * \lambda)(y) \\
&= \int f\, d\lambda.
\end{aligned}
$$

□

Stationary probability measures are interesting for a variety of reasons, as we will see. One of these is that integration against a stationary measure yields harmonic functions. This connection will be explored in depth in Chapter 12

Proposition 11.2.5 *If λ is a μ-stationary probability measure for a Γ-action on a compact metric space $\mathcal{Y}$, then for any bounded, measurable function $f : \mathcal{Y} \to \mathbb{R}$ the function $u : \Gamma \to \mathbb{R}$ defined by*

$$
u(g) = \int_{\mathcal{Y}} f(g \cdot y)\, d\lambda(y) = \int_{\mathcal{Y}} f(y)d(g \cdot \lambda)(y) \tag{11.2.4}
$$

is harmonic.

Proof. Relations (11.2.2) and (11.2.3) imply that for any $g \in \Gamma$,

$$
\begin{aligned}
u(g) &= \int f(g \cdot y)\, d\lambda(y) \\
&= \int f(g \cdot y)\, d(\mu * \lambda)(y) \\
&= \sum_{x\in\Gamma} \mu(x) \int f(g \cdot y)\, d(x \cdot \lambda)(y) \\
&= \sum_{x\in\Gamma} \mu(x) \int f(g \cdot (x \cdot y))\, d\lambda(y) \\
&= \sum_{x\in\Gamma} \mu(x) \int f(gx \cdot y)\, d\lambda(y) \\
&= \sum_{x\in\Gamma} \mu(x) u(gx).
\end{aligned}
$$

□

Stationary probability measures need not be unique, nor need the support of a stationary measure be the entire Γ-space. The trivial action $\Phi(g)$ =identity

on a space $\mathcal{Y}$ with infinitely many points, for instance, has infinitely many μ-stationary measures for any probability measure μ on Γ; these include all of the unit point masses $\{\delta_y\}_{y \in \mathcal{Y}}$. Such actions are of no interest, other than as a source of counterexamples. In Section 11.3 we will show that for *transitive* group actions, μ-stationary measures cannot have atoms.

Definition 11.2.6 Let $\mathcal{Y}$ be a compact Γ-space and μ a probability measure on Γ. A Borel probability measure λ on $\mathcal{Y}$ is Γ-*invariant* (or simply *invariant*, if there is no ambiguity about the group Γ) if

$$\lambda = g \cdot \lambda \quad \text{forevery } g \in \Gamma. \tag{11.2.5}$$

If λ is invariant, then it must also be μ-stationary for every probability measure μ on Γ. But although μ-stationary measures always exist, Γ-invariant measures need not.

Proposition 11.2.7 *A finitely generated group Γ is amenable only if every compact Γ-space admits a Γ-invariant Borel probability measure.*

Remark 11.2.8 The converse is also true: if every compact Γ-space admits a Γ-invariant Borel probability measure then Γ is amenable. We shall prove this in Section 12.4 (see Proposition 12.4.4), using the existence theorem for Furstenberg-Poisson boundaries.

Proof. This uses a variation of the Krylov-Bogolyubov argument. Suppose that Γ is amenable, and let $\Phi : \Gamma \to \mathrm{Homeo}(\mathcal{Y})$ be a Γ-action on a compact metric space $\mathcal{Y}$. Because Γ is amenable, there exist finite sets $F_n \subset \Gamma$ such that

$$\lim_{n\to\infty} \frac{|\partial F_n|}{|F_n|} = 0. \tag{11.2.6}$$

Fix any point $y \in \mathcal{Y}$, and for each $n = 1, 2, \cdots$, define a Borel probability measure λ_n by averaging the translates of δ_y by elements of the set F_n, that is,

$$\lambda_n = \frac{1}{|F_n|} \sum_{x \in F_n} x \cdot \delta_y = \frac{1}{|F_n|} \sum_{x \in F_n} \delta_{x \cdot y},$$

where δ_z denotes the unit point mass at z. By the Helly Selection Principle (see Section A.7 of the Appendix), the sequence λ_n contains a weakly convergent subsequence λ_m. Let λ be the (weak) limit of this subsequence. $\square$

Exercise 11.2.9 Finish the proof by showing that the measure λ is Γ-invariant.

HINT: It suffices to prove that for any continuous function $f : \mathcal{Y} \to \mathbb{R}$ and each generator $a \in \mathbb{A}$,

$$\int f \, d\lambda = \int f \, d(a \cdot \lambda).$$

For this, show that for any $n \in \mathbb{N}$,

$$\left|\int f\, d\lambda_n - \int f\, d(a \cdot \lambda_n)\right| \leq 2\, \|f\|_\infty \frac{|\partial F_n|}{|F_n|}$$

and then use the definition of λ as a weak limit.

Exercise 11.2.10 † Let Γ be an infinite, finitely generated group and $\mathcal{Y}$ a Γ-space. Let σ be the forward shift operator on the infinite product space Γ^∞ and P a σ-invariant Borel probability measure on Γ^∞. Denote elements of Γ^∞ by $\mathbf{x} = (x_0, x_1, x_2, \cdots)$.

(A) Show that there exists a Borel probability measure Q on the product space $\Gamma^\infty \times \mathcal{Y}$ that is invariant for the transformation

$$T(\mathbf{x}, y) = (\sigma\mathbf{x}, x_0 \cdot y)$$

and has marginal distribution P on Γ^∞, that is, for any Borel set $B \subset \Gamma^\infty$,

$$Q(B \times \mathcal{Y}) = P(B).$$

(B) Show by example that ergodicity of the measure P under σ does not imply ergodicity of the measure Q under T.

(C) Verify that if $P = \mu \times \mu \times \cdots$ is an infinite product measure, where μ is a probability distribution on Γ, then for any μ-stationary Borel probability measure λ on $\mathcal{Y}$ the product measure $P \times \lambda$ is invariant for the transformation T.

HINT: For (B), try the additive group $\Gamma = \mathbb{Z}$ and $\mathcal{Y} = S^1$ with the action $m \cdot \theta := \theta + m\pi \mod 2\pi$. The uniform distribution on S^1 (Lebesgue/2π) is μ-stationary for any probability distribution μ on $\mathbb{Z}$.

11.3 Transitive Group Actions

Definition 11.3.1 An action of a group Γ on a compact metric space $\mathcal{Y}$ by homeomorphisms is *transitive* if every orbit is dense, that is, if for every $y \in \mathcal{Y}$ the set $\Gamma \cdot y := \{g \cdot y \,:\, g \in \Gamma\}$ is a dense subset of $\mathcal{Y}$.

Exercise 11.3.2 Show that the natural action of the free group $\mathbb{F}_k$ on $\partial\mathbb{F}_k$ (see Example 11.1.2) is transitive.

Exercise 11.3.3 This exercise outlines a proof that the action (11.1.1) of $SL(2, \mathbb{Z})$ on the unit circle S^1 is transitive. For brevity, write $\Gamma = SL(2, \mathbb{Z})$. Denote by

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

the standard unit vectors in $\mathbb{R}^2$.

(A) Show that for each $i = 1, 2$ the set $\{g\mathbf{e}_i\}_{g\in\Gamma}$ (here gv denotes ordinary matrix multiplication) is the set of all nonzero vectors v whose entries are relatively prime integers.
HINTS: The fact that the entries of any $g\mathbf{e}_i$ in the orbit are relatively prime integers follows because the determinant of any $g \in \Gamma$ is 1. To show that all such vectors occur in the orbit, use the fact that for any two relatively prime integers a, b there exist integers c, d such that $ad - bc = 1$.
(B) Conclude that for each $i = 1, 2$ the orbit $\Gamma \cdot \mathbf{e}_i$ under the action (11.1.1) is dense in S^1.
(C) Fix $v \in S^1$. Show that if there exists a sequence of elements $g_n \in \Gamma$ such that $\lim_{n\to\infty} g_n \cdot v = \mathbf{e}_1$ then the orbit $\Gamma \cdot v$ is dense in S^1.
(D) Now show that for any $v \in S^1$ there exists a sequence of elements $g_n \in \Gamma$ such that $\lim_{n\to\infty} g_n \cdot v = \mathbf{e}_1$.

HINT: First show that it suffices to prove this for unit vectors v in the first quadrant. Then use the matrices

$$g_n = \begin{pmatrix} n & n-1 \\ 1 & 1 \end{pmatrix}.$$

Proposition 11.3.4 *Suppose that Γ acts transitively on a compact, metrizable space $\mathcal{Y}$ with infinitely many points. If μ is the step distribution of an irreducible random walk on Γ, then each μ-stationary probability measure λ on $\mathcal{Y}$ is nonatomic (that is, there is no singleton $\{y\} \subset \mathcal{Y}$ such that $\lambda(\{y\}) > 0$) and assigns positive probability to every nonempty open subset of $\mathcal{Y}$.*

Proof. An *atom* of a Borel measure λ on topological space $\mathcal{Y}$ is a point $y \in \mathcal{Y}$ such that λ attaches positive mass to the singleton $\{y\}$. Obviously, for any Borel probability measure the number of atoms of size at least $1/m$ cannot exceed m. Thus, if λ has an atom then there must be an atom y_* of maximal mass $\lambda(y_*) > 0$. If λ is μ-stationary then for every $n \geq 1$,

$$\lambda(y_*) = \sum_{g\in\Gamma} \mu^{*n}(g)\lambda(g^{-1} \cdot y_*),$$

and so every point $g^{-1} \cdot y_*$ such that $\mu^{*n}(g) > 0$ for some $n \in \mathbb{N}$ must also be an atom of size $\lambda(y_*)$. Since μ is the step distribution of an irreducible random walk, for every $g \in \Gamma$ there exists $n \in \mathbb{N}$ such that $\mu^{*n}(g) > 0$; consequently, every point of the orbit $\Gamma \cdot y_*$ must be an atom of size $\lambda(y_*)$. Our hypotheses imply that the orbit $\Gamma \cdot y_*$ is infinite; this contradicts the fact that the number of atoms of a size $\geq 1/m$ cannot exceed m. Therefore, λ must be nonatomic.

Any Borel probability measure ν on a compact metric space must have a *point of density*, that is, a point y such that ν assigns positive probability to every open ball

of positive radius centered at y. (Exercise: Why?) Let $y \in \mathcal{Y}$ be a point of density for λ, and let $J \subset \mathcal{Y}$ be a nonempty open set. By hypothesis, the orbit $\Gamma \cdot y$ is dense in $\mathcal{Y}$, so there exist $g_* \in \Gamma$ and an open set I containing the point y such that $g_* \cdot I \subset J$. Since λ is μ-stationary, for any $n \in \mathbb{N}$ we have

$$\lambda(J) = \sum_{g \in \Gamma} \mu^{*n}(g)\lambda(g^{-1} \cdot J) \geq \mu^{*n}(g_*)\lambda(I).$$

It follows that $\lambda(J) > 0$, because irreducibility implies that $\mu^{*n}(g_*) > 0$ for some $n \in \mathbb{N}$. □

Exercise 11.3.5 Suppose that Γ acts transitively (by permutations) on a finite set $\mathcal{Y}$. Show that if μ is the step distribution of an irreducible random walk on Γ then the uniform distribution on $\mathcal{Y}$ is the unique μ-stationary distribution on $\mathcal{Y}$.

11.4 μ-Processes and μ-Boundaries

The Cayley graphs of some finitely generated groups Γ admit compactifications $\Gamma \cup \partial\Gamma$ such that the natural action of Γ on the Cayley graph (by left multiplication) extends to an action on $\Gamma \cup \partial\Gamma$. This is the case, as we have seen, for the free groups $\mathbb{F}_k$ on $k \geq 2$ generators, and as we will see in Chapter 13, for the much larger class of *hyperbolic groups*. In some such cases, random walk paths $(X_n)_{n \geq 0}$ will converge almost surely to points Z of the boundaries $\partial\Gamma$. The Markov property implies that if this is the case, then the distribution of the limit Z is μ-stationary, where μ is the step distribution of the random walk. The key property of such a random variable Z — which is the focus of this section — is that it is a measurable, equivariant function of the random walk path, in the sense of Remark 11.4.3 to follow.

Assumption 11.4.1 *Assume throughout this section that $\mathcal{Y}$ is a compact Γ-space and that $(X_n)_{n \geq 0}$ is a transient, irreducible random walk on Γ with increments $\xi_n = X_{n-1}^{-1} X_n$ and step distribution μ.*

Definition 11.4.2 A measurable mapping $\zeta : \Gamma^\infty \to \mathcal{Y}$ is a *boundary map* if

$$\zeta(\xi_1, \xi_2, \cdots) = \xi_1 \cdot \zeta(\xi_2, \xi_3, \cdots) \quad \text{almost surely.} \tag{11.4.1}$$

Remark 11.4.3 The existence of a boundary map ζ is equivalent to the existence of an *equivariant boundary map* Y for the group action, that is, a shift-invariant (and therefore $\mathcal{I}$-measurable) transformation $Y : \Gamma^\infty \to \mathcal{Y}$ such that for all $x, g \in \Gamma$,

$$Y(gX_0, gX_1, gX_2, \cdots) = g \cdot Y(X_0, X_1, X_2, \cdots) \quad P^x\text{−almostsurely.} \tag{11.4.2}$$

The correspondence is given by

$$g \cdot \zeta(g_1, g_2, \cdots) = Y(g, gg_1, gg_1g_2, \cdots). \tag{11.4.3}$$

It is routine to check that if Ysatisfies the equivariance condition (11.4.3), then the function ζ defined by (11.4.3) is a boundary map, and conversely, that if ζ is a boundary map then the function Y defined by (11.4.3) is $\mathcal{I}$-measurable and satisfies (11.4.2).

Lemma 11.4.4 *If $\zeta : \Gamma^\infty \to \mathcal{Y}$ is a boundary map then the distribution λ of the $\mathcal{Y}$-valued random variable $Z = \zeta(\xi_1, \xi_2, \cdots)$ under the measure $P = P^1$ is μ-stationary.*

Proof. Since the increments of the random walk are i.i.d., the random variables ξ_1 and $Z' := \zeta(\xi_2, \xi_3, \cdots)$ are independent, and Z' has the same distribution λ as does Z. Consequently, the random variable $\xi_1 \cdot Z'$ has distribution $\mu * \lambda$, by definition of the convolution operation (equation (11.2.3)). Now if ζ is a boundary map, then $Z = \xi_1 \cdot Z'$, and so it follows that $\lambda = \mu * \lambda$. □

Exercise 11.4.5 Let $\Gamma = \mathbb{F}_k$ be the free group on k generators (for some $k \geq 2$) and let Z be the almost sure limit of the path $(X_n)_{n\geq 0}$ with respect to the metric (1.6.6) (see Exercise 1.6.6). Prove that $Z = \zeta(\xi_1, \xi_2, \cdots)$ for some boundary map ζ.

Definition 11.4.6 Let λ be a μ-stationary probability measure on $\mathcal{Y}$. The pair $(\mathcal{Y}, \lambda)$ is a *μ-boundary* if there is a boundary mapping $\zeta : \Gamma^\infty \to \mathcal{Y}$ such that the random variable $Z = \zeta(\xi_1, \xi_2, \cdots)$ has distribution λ.

Proposition 11.4.7 *Let $\zeta : \Gamma^\infty \to \mathcal{Y}$ be a boundary mapping for an irreducible random walk $(X_n)_{n\geq 0}$ with increments $\xi_1, \xi_2, \cdots$, and for each $n \geq 0$ set*

$$Z_n = \zeta(\xi_{n+1}, \xi_{n+2}, \cdots). \tag{11.4.4}$$

*Then the sequence of random vectors (ξ_n, Z_n) (called the μ-*process *associated with ζ) is stationary and ergodic.*

Proof. We may assume that the underlying probability space is the sequence space $\Omega = \Gamma^\infty$ with product measure $P = \mu \times \mu \times \cdots$, because stationarity and ergodicity of an infinite sequence are properties of its joint distribution (cf. Definition 2.1.5). The relevant measure-preserving transformation is the forward shift operator $\sigma : \Gamma^\infty \to \Gamma^\infty$. The terms (ξ_n, Z_n) of the sequence can be expressed as

$$(\xi_n, Z_n) = (\omega_1, \zeta \circ \sigma) \circ \sigma^{n-1}$$

where $\omega_1 : \Gamma^\infty \to \Gamma$ is the first coordinate evaluation map. Since any product measure is σ-invariant and ergodic (by Proposition 2.1.3), the assertion follows. □

The significance of Proposition 11.4.7 is that it provides an *ergodic* stationary sequence (ξ_n, Z_n), to which the conclusion of Birkhoff's Theorem 2.1.6 applies. There is a catch, of course: a μ-process requires a boundary mapping. Do these always exist? Unfortunately, they do not, even for transitive group actions. See

Exercise 11.4.10 below for an example. The geometric obstruction in this example, as you will see, is the "rigidity" of the group action. In contrast, the actions of Examples 11.1.2 and 11.1.3 are anything but rigid: as we will prove later, in each of these examples, for any irreducible random walk X_n on the group there is a nonatomic probability measure λ on the Γ-space $\mathcal{Y}$ such that with probability one the measures $X_n \cdot \lambda$ collapse to point masses as $n \to \infty$. In fact, this is a *necessary* condition, as we will show in Proposition 11.4.9. The following lemma shows that for a μ-stationary measure λ, the (random) measures $X_n \cdot \lambda$ always converge weakly to something.

Lemma 11.4.8 *For any μ-stationary probability measure λ on $\mathcal{Y}$ and every $x \in \Gamma$, with P^x-probability one, the sequence $(X_n \cdot \lambda)_{n \geq 0}$ converges weakly.*

Proof. It suffices to prove this for $x = 1$. Because the space $\mathcal{Y}$ is compact, Helly's Selection Principle implies that every sequence of Borel probability measures on $\mathcal{Y}$ has a weakly convergent subsequence. Thus, it suffices to show that with P-probability one, the sequence $(X_n \cdot \lambda)_{n \geq 0}$ has at most one weak limit point, and for this it is enough (cf. Exercise A.7.7 of the Appendix) to prove that for each real-valued continuous function $f \in C(\mathcal{Y})$ the sequence $(u(X_n))_{n \geq 0}$ converges almost surely, where $u : \Gamma \to \mathbb{R}$ is the function defined by

$$u(x) = \int f \, d(x \cdot \lambda) = \int_{y \in \mathcal{Y}} f(x \cdot y) \, d\lambda(y). \tag{11.4.5}$$

By Proposition 11.2.5, the function u is harmonic. Moreover, since $\mathcal{Y}$ is compact any continuous function $f : \mathcal{Y} \to \mathbb{R}$ is bounded, and so u is a *bounded* harmonic function. The almost sure convergence of the sequence $(u(X_n))_{n \geq 0}$ is therefore a consequence of the Martingale Convergence Theorem (cf. Theorem 8.0.1). □

Proposition 11.4.9 *There exists a boundary map ζ for the random walk $(X_n)_{n \geq 0}$ if and only if there is a μ-stationary probability measure λ such that the sequence $(X_n \cdot \lambda)_{n \geq 0}$ converges weakly, with P-probability one, to a point mass δ_Z at a random point Z. If this is the case then ζ is the Borel measurable function of the increment sequence $(\xi_n)_{n \geq 1}$ such that*

$$Z = \zeta(\xi_1, \xi_2, \cdots). \tag{11.4.6}$$

Proof. ($\Longleftarrow$) The weak convergence condition can be restated as follows: for almost every point $\omega \in \Omega$ of the underlying probability space there exists a point $Z(\omega) \in \mathcal{Y}$ such that for every continuous function $f : \mathcal{Y} \to \mathbb{R}$,

$$\lim_{n \to \infty} u(X_n(\omega)) = f(Z(\omega)) \tag{11.4.7}$$

where u is defined by equation (11.4.5). The limit $f(Z)$ must be measurable with respect to the σ-algebra $\mathcal{F}_\infty := \sigma((\xi_n))_{n \geq 0}$, since each $u(X_n)$ is; since the location of the point Z is determined by countably many functions $f \in C(\mathcal{Y})$

(cf. Corollary A.7.6), it follows that Z is measurable with respect to $\mathcal{F}_\infty$. Thus, $Z = \zeta(\xi_1, \xi_2, \cdots)$ for some Borel measurable function $\zeta : \Gamma^\infty \to \mathcal{Y}$.

It remains to show that the function ζ satisfies the equivariance requirement (11.4.1). Let $(\tilde{X}_n)_{n\geq 0}$ be the random walk with increments $\xi_2, \xi_3, \cdots$, that is, $\tilde{X}_n = \xi_2\xi_3\cdots\xi_{n+1}$, and let $\tilde{Z} = Z_1 = \zeta(\xi_2, \xi_3, \cdots)$. Since both sequences $(X_n)_{n\geq 0}$ and $(\tilde{X}_n)_{n\geq 0}$ are random walks with step distribution μ, the hypothesis (cf. equation (11.4.7)) implies that with probability one,

$$\delta_{\tilde{Z}} = \text{weak}-\lim_{n\to\infty} \tilde{X}_n \cdot \lambda \quad \text{and}$$

$$\delta_Z = \text{weak}-\lim_{n\to\infty} X_n \cdot \lambda = \text{weak}-\lim_{n\to\infty} \xi_1 \cdot (\tilde{X}_n \cdot \lambda).$$

But by Exercise 11.2.3,

$$\text{weak}-\lim_{n\to\infty} \xi_1 \cdot (\tilde{X}_n \cdot \lambda) = \xi_1 \cdot \text{weak}-\lim_{n\to\infty} \tilde{X}_n \cdot \lambda = \xi_1 \cdot \delta_{\tilde{Z}} = \delta_{\xi_1 \cdot \tilde{Z}},$$

so it follows that $Z = \xi_1 \cdot \tilde{Z}$ almost surely. This proves equation (11.4.1). □

Proof. ($\Longrightarrow$) If $\zeta : \Gamma^\infty \to \mathcal{Y}$ is a boundary map, then the distribution λ of $Z := \zeta(\xi_1, \xi_2, \cdots)$ is μ-stationary, by Lemma 11.4.4, so for each $n \in \mathbb{N}$ the random variable $Z_n = \zeta(\xi_{n+1}, \xi_{n+2}, \cdots)$ has distribution λ and is independent of the σ-algebra $\mathcal{F}_n = \sigma(\xi_i)_{1\leq i\leq n}$. Since ζ is a boundary map, $Z = X_n \cdot Z_n$ almost surely for every $n \in \mathbb{N}$. Consequently, for any continuous function $f : \mathcal{Y} \to \mathbb{R}$,

$$\int f \, d(X_n \cdot \lambda) = E(f(X_n \cdot Z_n) \,|\, \mathcal{F}_n) = E(f(Z) \,|\, \mathcal{F}_n).$$

But the random variable $f(Z)$ is bounded and measurable with respect to the σ-algebra $\mathcal{F}_\infty$ generated by $\xi_1, \xi_2, \cdots$, so the Martingale Convergence Theorem implies that $\lim_{n\to\infty} E(f(Z) \,|\, \mathcal{F}_n) = f(Z)$ almost surely. □

Exercise 11.4.10 Let $\Gamma = \mathbb{Z}$ and $\mathcal{Y} = S^1 = \{z \in \mathbb{C} : |z| = 1\}$. Fix an irrational number $0 < \theta < 1$ and consider the $\mathbb{Z}$-action

$$m \cdot z = e^{2\pi i m\theta} z. \tag{11.4.8}$$

(A) Prove that If μ is the step distribution of an irreducible random walk then the only μ-stationary measure on $\mathcal{Y}$ is the uniform distribution (Lebesgue/2π).
HINT. One way to do this is by Fourier analysis: show that the Fourier transform of a stationary measure λ satisfies

$$\hat{\lambda}(n) = \delta_0(n) \quad \text{for all } n \in \mathbb{Z},$$

where δ_0 is the Kronecker delta function.

(B) Show that no irreducible random walk on $\mathbb{Z}$ admits a boundary map for the action (11.4.8).

Corollary 11.4.11 *If μ is the step distribution of an irreducible random walk $(X_n)_{n\geq 0}$ on $SL(2,\mathbb{Z})$ then there is a unique μ-stationary probability measure λ on $\mathbb{RP}^1$ (relative to the natural action (11.1.1)), and the pair $(\mathbb{RP}^1, \lambda)$ is a μ-boundary.*

Proof. By Exercise 11.3.3, the group action is transitive, so any μ-stationary probability measure is nonatomic and assigns positive probability to every nonempty open arc of $\mathbb{RP}^1$. By Lemma 11.4.8, with probability one the measures $X_n \cdot \lambda$ converge weakly; thus, by Proposition 11.4.9, to prove the existence of a boundary map it is enough to show that the only possible weak limits are point masses. This is a consequence of the following lemma, because the transience of the random walk implies that $\|X_n\| \to \infty$ almost surely as $n \to \infty$, where $\|\cdot\|$ denotes the usual matrix norm. (Explanation: For any $C < \infty$ there are only finitely many elements of $SL(2,\mathbb{Z})$ with norm $\leq C$.) Given the existence of a boundary map, uniqueness of the stationary measure λ will follow from Proposition 11.4.13. □

Lemma 11.4.12 *For any sequence $(g_n)_{n\in\mathbb{N}}$ in $SL(2,\mathbb{R})$ such that*

$$\lim_{n\to\infty} \|g_n\| = \infty \tag{11.4.9}$$

there exists a subsequence $(g_m)_{m\in\mathbb{N}}$ such that

$$\lim_{m\to\infty} \frac{g_m}{\|g_m\|} = h \tag{11.4.10}$$

exists. Any such limit matrix h has rank one and norm 1. Consequently, for any nonatomic probability measure λ on $\mathbb{RP}^1$,

$$\text{weak-} \lim_{m\to\infty} g_m \cdot \lambda = \delta_{\pm v}, \tag{11.4.11}$$

where $v \in S^1$ is a basis for the column space of h.

Proof. For each $n \in \mathbb{N}$ the normalized matrix $\tilde{g}_n := g_n/\|g_n\|$ has norm 1, so its entries are bounded in absolute value by 1. Hence, by the Bolzano-Weierstrass Theorem, the sequence has a convergent subsequence. The limit matrix h must have determinant 0, because $\det(\tilde{g}_n) = 1/\|g_n\| \to 0$, and so h must have rank < 2. Moreover, since the matrix norm is continuous on the space of all real 2×2 matrices, h must have norm 1, and therefore also rank 1.

Let v and w be unit vectors whose linear spans are the column space and the null space of h. For every unit vector $u \neq \pm w$, the vector hu is a nonzero scalar multiple $hu = \alpha(u)v$ of v. Consequently, by (11.4.10) and the continuity of matrix multiplication,

$$\lim_{m\to\infty} \frac{g_m u}{\|g_m\|} = \alpha(u)v,$$

which implies that

$$\lim_{m\to\infty} g_m \cdot (\pm u) = \pm v. \tag{11.4.12}$$

Since this convergence holds for all unit vectors u except for the antipodal pair $\pm w$, the weak convergence (11.4.11) follows, by the hypothesis that λ is nonatomic: in particular, since $\lambda(\{\pm w\}) = 0$, (11.4.12) implies that for any continuous function $f : \mathbb{RP}^1 \to \mathbb{R}$,

$$\lim_{m\to\infty} \int f(g_m \cdot u)\, d\lambda(u) = f(\pm v),$$

by the Dominated Convergence Theorem. □

Proposition 11.4.13 *Let μ be the step distribution of an irreducible random walk $(X_n)_{n\geq 0}$ on a finitely generated group Γ, and let $(\mathcal{Y}, \lambda)$ be a μ-boundary with boundary map ζ. If the action of Γ on $\mathcal{Y}$ is transitive, then λ is the only μ-stationary probability measure, and furthermore, for any point $y \in \mathcal{Y}$,*

$$\lim_{n\to\infty} X_n \cdot y = Z = \zeta(\xi_1, \xi_2, \cdots) \quad \text{almostsurely.} \tag{11.4.13}$$

Proof. By Proposition 11.3.4, any μ-stationary measure is nonatomic and assigns positive probability to every nonempty open set. This, together with the hypothesis that $X_n \cdot \lambda \to \delta_Z$ weakly, implies that for any point $y' \in \mathcal{Y}$ and any positive numbers $r, \delta > 0$,

$$X_n \cdot \mathbb{B}_r(y')^c \subset \mathbb{B}_\delta(Z) \quad \text{eventually,} \tag{11.4.14}$$

with probability one. □

Exercise 11.4.14 Prove this.

HINT: Use the definition of weak convergence with a function f that has support in $\mathbb{B}_\delta(Z)$ and is identically 1 in $\mathbb{B}_{\delta/2}(Z)$.

If (11.4.14) holds for some $y' \in \mathcal{Y}$ and all $r, \delta > 0$ then it must be the case that with probability one, $\lim_{n\to\infty} X_n \cdot y = Z$ for every $y \neq y'$. Since y' is arbitrary, this proves (11.4.13).

Now let λ' be any μ-stationary probability measure on $\mathcal{Y}$ and Y a random variable with distribution λ' that is independent of the random walk $(X_n)_{n\geq 0}$. Stationarity implies that for each $n \in \mathbb{N}$ the random variable $X_n \cdot Y$ has distribution λ', while (11.4.13) implies that $\lim_{n\to\infty} X_n \cdot Y = Z$ almost surely. Consequently, the probability measure λ' must be the distribution of the random variable Z.

11.5 Boundaries and Speed

Furstenberg discovered (cf. [45] and [47]) that boundaries can carry important information about the speeds of random walks. In this section, we shall look at two elementary but instructive cases: free groups and the matrix group $SL(2, \mathbb{Z})$. In Sections 11.6 and 11.7, we will show that for any finitely generated group Γ there is a Γ-space, called the *Busemann boundary*, that encapsulates all of the important information about speeds of random walks on Γ.

11.5.1 A. Random Walks on $\mathbb{F}_k$

We have seen that for any probability measure μ on $\mathbb{F}_k$, if the random walk $(X_n)_{n\geq 0}$ with step distribution μ is irreducible then $\partial\mathbb{F}_k$ is a μ-boundary. Consequently, the stationary measure λ is unique, and is the distribution of the random variable

$$Z := \text{a.s.} - \lim_{n\to\infty} X_n, \tag{11.5.1}$$

where the limit is the end of the Cayley graph to which the trajectory $(X_n)_{n\geq 0}$ converges. The key to relating the convergence (11.5.1) with the speed of the random walk is that multiplication of an infinite word ω by a group element g shifts the "tail" of ω by an integer distance $B(g;\omega)$ equal to the word length $|g|$ of g minus twice the length of the initial prefix of ω cancelled in the reduction of the word $g\cdot\omega$. Thus, for all sufficiently large $n \in \mathbb{N}$,

$$(g\cdot\omega)_n = \omega_{n+B(g;\omega)}. \tag{11.5.2}$$

For example, if $g = aba$ and $\omega = \bar{a}bb\bar{a}\bar{a}\cdots$ (where $\bar{a} = a^{-1}$ and $\bar{b} = b^{-1}$) then $|g| = 3$ and only the initial letter $\bar{a}$ is cancelled in the reduction, so $B(g;\omega) = 1$:

$$\omega = \bar{a}bb\bar{a}\bar{a}\cdots,$$
$$g\cdot\omega = abbb\bar{a}\bar{a}\cdots.$$

Exercise 11.5.1 Check that the mapping B has the following *cocycle property*: for any $\omega \in \partial\mathbb{F}_k$ and any two group elements g_1, g_2,

$$B(g_2g_1;\omega) = B(g_1;\omega) + B(g_2; g_1\cdot\omega). \tag{11.5.3}$$

Theorem 11.5.2 (Furstenberg [48]) *Let $(X_n)_{n\geq 0}$ be an irreducible random walk on the free group $\mathbb{F}_k$ (for some $k \geq 2$) whose step distribution μ has finite first moment $\sum_{g\in\Gamma} |g|\mu(g)$, where $|\cdot|$ denotes the word length norm. Let λ be the exit distribution of the random walk, equivalently, the unique μ-stationary probability*

measure on $\partial\mathbb{F}_k$ for the natural $\mathbb{F}_k$-action. Then the speed of the random walk is given by

$$\ell = \sum_{g \in \mathbb{F}_k} \mu(g) \int B(g; \omega)\, d\lambda(\omega). \tag{11.5.4}$$

Proof. Recall (see Example 3.2.6 and Corollary 3.3.2) that the *speed* of a random walk $X_n = \xi_1 \xi_2 \cdots \xi_n$ with independent, identically distributed steps ξ_i such that $E|\xi_1| < \infty$ is the limit

$$\ell = \lim_{n\to\infty} \frac{E|X_n|}{n} = \text{a.s.}- \lim_{n\to\infty} \frac{|X_n|}{n}.$$

Since the steps ξ_i are i.i.d., for any $n \in \mathbb{N}$ the distribution of the product $X_n = \xi_1 \xi_2 \cdots \xi_n$ is the same as that of the reverse product $\xi_n \xi_{n-1} \cdots \xi_1$; consequently (see Remark 3.2.7), the speed of the random walk $\tilde{X}_n := \xi_1^{-1} \xi_2^{-1} \cdots \xi_n^{-1}$ with increments ξ_i^{-1} is also ℓ. Thus,

$$\ell = \lim_{n\to\infty} \frac{E|\tilde{X}_n|}{n} = \text{a.s.}- \lim_{n\to\infty} \frac{|\tilde{X}_n|}{n}.$$

By Proposition 11.4.13, the space $\partial\mathbb{F}_k$ supports a unique μ-stationary distribution λ; by Exercise 11.4.5, this is the distribution of the random variable $Z = \text{a.s.}- \lim_{n\to\infty} X_n$. Exercise 11.4.5 also implies that $Z = \zeta(\xi_1, \xi_2, \cdots)$ where ζ is a boundary map. Set $Z_n = \zeta(\xi_{n+1}, \xi_{n+2}, \cdots)$; then by Proposition 11.4.7, the sequence $(\xi_n, Z_n)_{n \geq 0}$ is stationary and ergodic, and so by the Ergodic Theorem (Theorem 2.1.6)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} B(\xi_i; Z_{i-1}) = EB(\xi_1; Z_0) = \sum_{g \in \Gamma} \mu(g) \int B(g; \omega)\, d\lambda(\omega) \tag{11.5.5}$$

almost surely.

The cocycle property (11.5.3) implies that for every $n \in \mathbb{N}$,

$$B(\tilde{X}_n^{-1}; Z_0) = B(\xi_n \xi_{n-1} \cdots \xi_1; Z_0) = \sum_{i=1}^{n} B(\xi_i; Z_{i-1}).$$

By definition, $B(\tilde{X}_n^{-1}; Z_0) = |\tilde{X}_n^{-1}| - 2C(\tilde{X}_n^{-1}; Z_0)$, where $C(g; \omega)$ is the length of the initial prefix of ω that is cancelled in the reduction of the infinite word $g \cdot \omega$. We claim that the limit $C_\infty := \limsup_{n\to\infty} C(\tilde{X}_n^{-1}; Z_0)$ is finite with probability 1; this will imply that with probability one,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}B(\xi_i; Z_{i-1}) = \lim_{n\to\infty}\frac{|\tilde{X}_n|}{n} = \ell,$$

which together with (11.5.5) will prove the equality (11.5.4).

It remains to prove the claim that $\limsup_{n\to\infty} C(\tilde{X}_n^{-1}; Z_0) < \infty$ almost surely. By definition, $C(\tilde{X}_n^{-1}; Z_0)$ is the length of the maximal common prefix of the infinite word Z_0 and the reduced word representing $\tilde{X}_n$; consequently, $\lim_{n\to\infty} C(\tilde{X}_n^{-1}; Z_0) \leq L$, where $L \leq \infty$ is the maximum integer (or $+\infty$) such that the random walk $(\tilde{X}_n)_{n\geq 0}$ visits the group element at distance L along the geodesic ray in the Cayley graph from 1 to $Z_0 \in \partial\mathbb{F}_k$. But the random walk $(\tilde{X}_n)_{n\geq 0}$ is irreducible, so by Proposition 11.3.4 its exit distribution is nonatomic; therefore,

$$P\left\{\lim_{n\to\infty}\tilde{X}_n = Z_0\right\} = 0.$$

Thus, with probability one, the random variable L is finite, and so the sequence $C(\tilde{X}_n^{-1}; Z_0)$ is bounded above. □

11.5.2 B. Random Walks on $SL(2,\mathbb{Z})$

Let μ be the step distribution of an irreducible random walk $(X_n)_{n\geq 0}$ on $\Gamma = SL(2,\mathbb{Z})$ with increments $\xi_n = X_{n-1}^{-1}X_n$, and assume that

$$E\log\|X_1\| = \sum_{g\in\Gamma}\mu(g)\log\|g\| < \infty. \tag{11.5.6}$$

According to the Furstenberg-Kesten Theorem (Corollary 3.3.4), the limit

$$\alpha := \lim_{n\to\infty}\frac{\log\|X_n\|}{n} \tag{11.5.7}$$

exists with probability one and is constant. This constant is the *speed* of the random walk relative to the (log-)matrix norm metric on $SL(2,\mathbb{Z})$; it is necessarily nonnegative, as elements of $SL(2,\mathbb{Z})$ have norms ≥ 1.

Exercise 11.5.3 Prove that the speed α of an irreducible random walk on $SL(2,\mathbb{Z})$ is strictly positive.

HINT: Since Γ is nonamenable, any irreducible random walk on Γ must have spectral radius less than 1, by Theorem 5.1.6 (or, for symmetric random walks, by Kesten's Theorem 5.1.5). Consequently, the Avez entropy cannot be 0. Now use the fact that for some $C < \infty$ the number of distinct elements of Γ with norm $\leq R$ is bounded by CR^4 for all $R > 0$.

Theorem 11.5.4 *The speed α of an irreducible random walk $(X_n)_{n\geq 0}$ with step distribution μ is given by the formula*

$$\alpha = \sum_{g \in SL(2,\mathbb{Z})} \mu(g) \int_{v \in \mathbb{RP}^1} \log \|gv\| \, d\lambda(v), \tag{11.5.8}$$

where λ is the unique μ-stationary probability measure λ for the natural action (11.1.1) of $SL(2,\mathbb{Z})$ on $\mathbb{RP}^1$. Furthermore, there exists a random variable $Z = \pm U = \zeta(\xi_1, \xi_2, \cdots)$ taking values in $\mathbb{RP}^1$ such that for every nonzero vector $v \in \mathbb{R}^2$, with probability one,

$$\lim_{n\to\infty} \frac{\|v^T X_n\|}{\|X_n\|} = |\langle v, U\rangle| \quad \text{and} \tag{11.5.9}$$

$$\lim_{n\to\infty} \pm \frac{X_n v}{\|X_n v\|} = Z \tag{11.5.10}$$

Here v^T denotes the transpose of v, and $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product; clearly, the value $|\langle v, U\rangle|$ does not depend on which of the two unit vectors U in the equivalence class Z is chosen.

Lemma 11.5.5 *There exists $\delta > 0$ with the following property: for every $g \in SL(2,\mathbb{R})$ there is an arc $J \subset S^1$ of length at least δ such that for every $u \in J$,*

$$\|gu\| \geq \frac{1}{2} \|g\|. \tag{11.5.11}$$

Proof. A real 2×2 matrix of determinant 1 maps the unit circle to an ellipse whose semi-major and semi-minor axes have lengths that multiply to 1, and the inverse matrix maps the unit vectors in these directions to the semi-major and semi-minor axes of another ellipse of the same area. Hence, there exist pairs u_+, u_- and v_+, v_- of orthogonal unit vectors and a scalar $\alpha = \|g\|$ such that

$$g = \alpha u_+ v_+^T + \alpha^{-1} u_- v_-^T.$$

Let w be any unit vector such that $|v_-^T w|^2 \leq \frac{1}{16}$; then $|v_+^T w|^2 \geq \frac{15}{16} \geq \frac{9}{16}$, and so

$$\|gw\| \geq \frac{3}{4}\alpha - \frac{1}{4}\alpha^{-1} \geq \frac{1}{2}\alpha = \frac{1}{2}\|g\|.$$

For any unit vector v_-, the set of unit vectors w such that $|v_-^T w|^2 \leq \frac{1}{16}$ is an arc whose length δ does not depend on v_-, so the lemma follows. □

Proof of Theorem 11.5.4. By Corollary 11.4.11 and Proposition 11.4.13, the μ-stationary probability measure λ on $\mathbb{RP}^1$ is unique, nonatomic, and charges every nonempty open arc of $\mathbb{RP}^1$. Moreover, the pair $(\mathbb{RP}^1, \lambda)$ is a μ-boundary, so there

exists a random variable $Z = \zeta(\xi_1, \xi_2, \cdots)$ with distribution λ such that with probability one

$$\text{weak} - \lim_{n\to\infty} X_n \cdot \lambda = \delta_Z.$$

Since the action (11.1.1) of $SL(2, \mathbb{Z})$ on $\mathbb{RP}^1$ is transitive, Proposition 11.4.13 implies that for every $\pm v \in \mathbb{RP}^1$.

$$\lim_{n\to\infty} X_n \cdot (\pm v) = Z \quad \text{almostsurely}.$$

The assertion (11.5.10) follows, because every nonzero vector $v \in \mathbb{R}^2$ is a scalar multiple of a unit vector.

The random walk $(X_n)_{n\geq 0}$ is transient, so with probability one, $\|X_n\| \to \infty$ as $n \to \infty$. Hence, by Lemma 11.4.12, every subsequence has a subsequence $(X_{n_m})_{m\in\mathbb{N}}$ such that the normalized matrices $X_{n_m}/\left\|X_{n_m}\right\|$ converge to a (random) rank one matrix H of norm 1, possibly depending on the subsequence $(X_{n_m})_{m\in\mathbb{N}}$. In view of (11.5.10), the column space of any such limit matrix H must be the line in the direction Z. Moreover, since H has norm one, it has a representation $H = UV^T$ with both U and V unit vectors; since the column space is the line in direction Z, we must have $Z = \pm U$. Consequently,

$$\lim_{m\to\infty} \frac{\left\|v^T X_{n_m}\right\|}{\left\|X_{n_m}\right\|} = \left\|v^T H\right\| = v^T U = |\langle v, U\rangle|.$$

Since this limit is independent of the subsequence, the relation (11.5.9) follows.

For each integer $n \geq 0$ define $Z_n = \zeta(\xi_{n+1}, \xi_{n+2}, \cdots)$. By Proposition 11.4.7, the sequence $(\xi_n, Z_n)_{n\geq 0}$ is stationary and ergodic, so the Ergodic Theorem implies that with probability one,

$$\beta = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log \|\xi_i Z_i\| = \lim_{n\to\infty} \frac{1}{n} \log \|X_n Z_n\|, \quad \text{where} \tag{11.5.12}$$

$$\beta = E \log \|\xi_1 Z_1\| = \sum_{g\in SL(2,\mathbb{Z})} \mu(g) \int_{v\in\mathbb{RP}^1} \log \|gv\| \, d\lambda(v). \tag{11.5.13}$$

Here $\|X_n Z_n\| = \|X_n U_n\|$ where U_n is either one of the two unit vectors such that $Z_n = \pm U_n$. Since $\|X_n U_n\| \leq \|X_n\|$, the Furstenberg-Kesten Theorem (11.5.7) implies that $\beta \leq \alpha$. To complete the proof of (11.5.8) we must show that $\alpha = \beta$.

The μ-stationary measure λ assigns positive probability to every nonempty open arc of $\mathbb{RP}^1$, so for any $\delta > 0$ there exists $\varepsilon > 0$ such that

$$\lambda(J) \geq \varepsilon \quad \text{for every arc of length } |J| = \delta.$$

Since each Z_n has distribution λ and is independent of X_n, it follows by Lemma 11.5.5 that for every $n \in \mathbb{N}$,

$$P\left\{\|X_n Z_n\| \geq \frac{1}{2}\|X_n\|\right\} \geq \varepsilon.$$

This, along with the limiting relations (11.5.7) and (11.5.12), implies that $\beta \geq \alpha$.

□

11.6 The Busemann Boundary

The theorems of Section 11.5 both depended on the existence of μ-boundaries $(\mathcal{Y}, \lambda)$ equipped with functions $B : \Gamma \times \mathcal{Y} \to \mathcal{R}$ that measure changes in distance brought about by left translations of the Cayley graph. In this section we will show that for every finitely generated group Γ there is a compact Γ-space $\partial_B \Gamma$, the *Busemann boundary*, that serves a similar purpose. In Section 11.7, we will use the Busemann boundary to obtain characterizations of the speed of a random walk analogous to those of Theorems 11.5.2 and 11.5.4.

A metric space $(\mathcal{Y}, d)$ is said to be *proper* if it has the *Heine-Borel* property, that is, every closed, bounded subset is compact. Any finitely generated group, when given its word metric, is proper. For any reference point $y_* \in \mathcal{Y}$ in a proper metric space $\mathcal{Y}$ there is an embedding $\Phi : \mathcal{Y} \to C(\mathcal{Y})$ in the space $C(\mathcal{Y})$ of continuous, real-valued functions on $\mathcal{Y}$ defined as follows: for each $y \in \mathcal{Y}$, let $\Phi_y \in C(\mathcal{Y})$ be the function

$$\Phi_y(z) := d(y, z) - d(y, y_*) \quad \text{for all } z \in \mathcal{Y}. \tag{11.6.1}$$

The function Φ_y takes the value 0 at the reference point y_*, and by the triangle inequality, for any points $y, y', z, z' \in \mathcal{Y}$

$$|\Phi_y(z') - \Phi_y(z)| \leq d(z, z'), \tag{11.6.2}$$

$$|\Phi_y(z) - \Phi_{y'}(z)| \leq 2d(y, y'), \quad \text{and} \tag{11.6.3}$$

$$|\Phi_y(z)| \leq d(z, y_*), \tag{11.6.4}$$

with equality in (11.6.4) if $z = y$. The first inequality (11.6.2) implies that for each $y \in \mathcal{Y}$ the function $\Phi_y \in C(\mathcal{Y})$ is 1-*Lipschitz*: by definition, a function $\Psi : \mathcal{Y} \to \mathcal{Z}$ between two metric spaces $(\mathcal{Y}, d)$ and $(\mathcal{Z}, d')$ is C-Lipschitz if for any two points $y, y' \in \mathcal{Y}$,

$$d'(\Psi(y), \Psi(y')) \leq C d(y, y'). \tag{11.6.5}$$

The second inequality (11.6.3) implies that the embedding $\Phi : \mathcal{Y} \to C(\mathcal{Y})$ is continuous relative to the metric ϱ on $C(\mathcal{Y})$ defined by

$$\varrho(f, g) := \sum_{n=1}^{\infty} 2^{-n} \sup_{y \in \mathbb{B}_n} |f(y) - g(y)|, \tag{11.6.6}$$

where $\mathbb{B}_n = \mathbb{B}_n(y_*)$ is the ball of radius n centered at the reference point y_*. (The topology induced by this metric is the topology of uniform convergence on compact sets.)

Lemma 11.6.1 *For any proper metric space $(\mathcal{Y}, d)$ the image $\Phi(\mathcal{Y})$ is sequentially pre-compact in the metric (11.6.6), that is, every sequence $(\Phi_{y_n})_{n\in\mathbb{N}}$ has a ϱ-convergent subsequence.*

Proof. We will prove this only for a special case of interest, where $\mathcal{Y} = \Gamma$ is a finitely generated group, d is the word metric for some finite, symmetric generating set, and the reference point y_* is the group identity. In this case the functions Φ_y are integer-valued, because the word metric only takes values in $\mathbb{Z}_+$, and for each $y \in \Gamma$ we have $\Phi_y(1) = 0$. Hence, by the Lipschitz property (11.6.2), for any points $y, z \in \Gamma$,

$$|\Phi_y(z)| \in \{-|z|, -|z| + 1, -|z| + 2, \cdots , |z|\},$$

where $|\cdot|$ is the word metric norm, and so by the Pigeonhole Principle, for any $z \in \Gamma$ and any sequence $(y_n)_{n\in\mathbb{N}}$ there is a subsequence $(y_i)_{i\in\mathbb{N}}$ along which the values $\Phi_{y_i}(z)$ are all equal. It follows, by Cantor's diagonal argument, that every sequence $(\Phi_{y_n})_{n\in\mathbb{N}}$ has a subsequence that stabilizes pointwise, and therefore converges in the metric ϱ. □

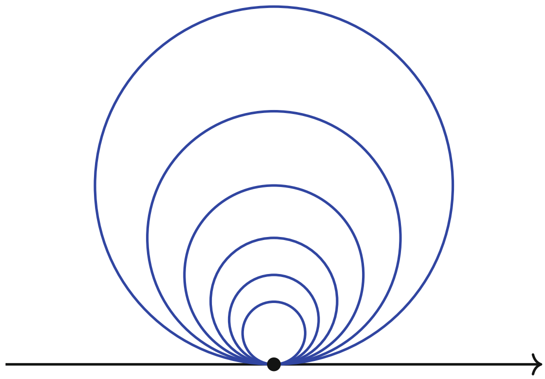
Exercise 11.6.2 Prove Lemma 11.6.1 in the general case, where $(\mathcal{Y}, d)$ is an arbitrary proper metric space.

HINT: Since the functions Φ_y are all 1-Lipschitz, it is enough to find subsequences that converge at each point z in a dense subset of $\mathcal{Y}$. Since the metric space $\mathcal{Y}$ is proper, it has a countable dense subset.

Definition 11.6.3 The *Busemann compactification* of a proper metric space $\mathcal{Y}$ is the closure $\mathcal{K} = \overline{\Phi(\mathcal{Y})}$ of the set $\Phi(\mathcal{Y}) = \{\Phi_y\}_{y\in\mathcal{Y}}$ in the metric space $(C(\mathcal{Y}), \varrho)$. The *Busemann boundary* $\partial_B\mathcal{Y}$ is the compact subset of $\mathcal{K}$ consisting of all functions h that arise as limits $h = \lim_{n\to\infty} \Phi_{y_n}$, where $(y_n)_{n\in\mathbb{N}}$ is a sequence such that $\lim_{n\to\infty} d(y_n, y_*) = \infty$. Elements of the Busemann boundary are called *horofunctions*.

Example 11.6.4 † The term *horofunction* originates in hyperbolic geometry. The *hyperbolic plane* can be viewed as the upper halfplane $\mathbb{H} = \{z \in \mathbb{C} : \Re z > 0\}$ with the Poincaré metric

Fig. 11.2 Horocycles



$$d_H(z_1, z_2) = \min_{\gamma \in \mathcal{P}(z_1,z_2)} \int_\gamma \frac{2\,|dz|}{\Im z}, \qquad (11.6.7)$$

where $\mathcal{P}(z_1, z_2)$ is the set of all piecewise differentiable paths from z_1 to z_2. See, for instance, Beardon [8] for background on hyperbolic geometry. The hyperbolic plane is of interest here because its isometry group is $SL(2, \mathbb{R})$, whose elements act on $\mathbb{H}$ by linear fractional transformation (cf. equation (11.1.4)). The Busemann boundary $\partial_B \mathbb{H}$ is the circle $\mathbb{R} \cup \{\infty\}$. For any real point $x \in \mathbb{R}$ on the boundary, the corresponding horofunction has as its level curves the circles tangent to the line $\mathbb{R}$ at x; these are called the *horocycles* based at x. See Figure 11.2.

Notational Convention: In the special case where $\mathcal{Y} = \Gamma$ is a finitely generated group endowed with the word metric d and reference point $y_* = 1$, the Busemann compactification will be denoted by $\overline{\Gamma}$ and the Busemann boundary by $\partial_B \Gamma$. In this case, the horofunctions are integer-valued functions that vanish at the group identity 1.

For certain groups Γ — including the *hyperbolic groups*, as we will show in Chapter 13 — the Busemann boundary $\partial_B \Gamma$ can be identified with the set of geodesic rays γ with initial point 1; the horofunction associated with γ is the function $h_\gamma : \Gamma \to \mathbb{Z}$ defined by

$$h_\gamma(g) := \lim_{n\to\infty} d(\gamma_n, g) - n. \qquad (11.6.8)$$

Here γ_n denotes the unique point on γ at distance n from the group identity. That the limit exists is an easy consequence of the triangle inequality.

Exercise 11.6.5 Show that for the free group $\Gamma = \mathbb{F}_k$ on $k \geq 2$ generators the Busemann compactification is homeomorphic to the space $\mathbb{T}_{2k} \cup \partial \mathbb{T}_{2k}$, where $\mathbb{T}_{2k}$ is the infinite $2k$-ary tree and $\partial \mathbb{T}_{2k}$ the space of ends (see Section 1.6), with metric d defined by (1.6.6).

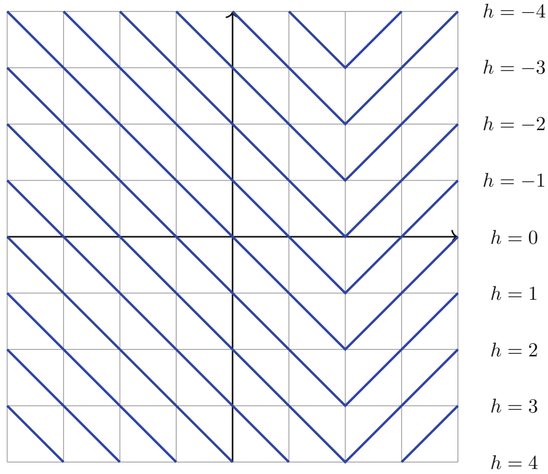


Fig. 11.3 A Horofunction on $\mathbb{Z}^2$

Exercise 11.6.6 Give a complete description of the Busemann compactification of the 2- and 3-dimensional integer lattices $\mathbb{Z}^2$, $\mathbb{Z}^3$, equipped with their word metrics.

HINT: Start with $d = 2$. In this case there are countably many horofunctions with "corners" (see Figure 11.3 for a depiction of the level sets of one of these) and four without corners.

Exercise 11.6.7 Show that when $\mathbb{Z}^d$ is equipped with the *Green metric* for the simple random walk (cf. Example 3.2.9), the Busemann compactification is the one-point compactification.

HINT: What does Proposition 9.6.7 tell you about the hitting probability function?

Note 11.6.8 When a finitely generated group Γ is equipped with the Green metric for an irreducible random walk, the corresponding Busemann boundary is known as the *Martin boundary* of the random walk. See [132], Chapter IV for more on Martin boundary theory for random walks, including a detailed discussion of the Martin representation of nonnegative harmonic functions.

Assume now that Γ acts by isometries on a proper metric space $(\mathcal{Y}, d)$. Fix a reference point $y_* \in \mathcal{Y}$ for the Busemann compactification $\mathcal{K} := \overline{\Phi(\mathcal{Y})}$. The action of Γ on $\mathcal{Y}$ induces corresponding actions on the Busemann compactification and boundary as follows: for any function $f \in C(\mathcal{Y})$ that vanishes at the reference point y_*, set

$$g \cdot f(z) = f(g^{-1} \cdot z) - f(g^{-1} \cdot y_*). \tag{11.6.9}$$

It is easily checked that this defines an action of Γ by homeomorphisms on the space of all continuous functions that vanish at y_*. To prove that its restriction to the Busemann boundary $\partial_B \mathcal{Y}$ is an action, we must show that the mapping (11.6.9)

sends horofunctions to horofunctions. For this it is enough to show that $\Gamma \cdot \Phi(\mathcal{Y}) \subset \Phi(\mathcal{Y})$, since this will imply that Γ maps limit points of $\Phi(\mathcal{Y})$ to other limit points. But this follows easily by the assumption that Γ acts on $\mathcal{Y}$ by *isometries*: in particular, for any $g \in \Gamma$ and any points $y, z \in \mathcal{Y}$

$$\begin{aligned} g \cdot \Phi_y(z) :&= \Phi_y(g^{-1} \cdot z) - \Phi_y(g^{-1} \cdot y_*) \\ &= d(y, g^{-1} \cdot z) - d(y, g^{-1} \cdot y_*) \\ &= d(g \cdot y, z) - d(g \cdot y, y_*) \\ &:= \Phi_{g \cdot y}(z). \end{aligned} \tag{11.6.10}$$

This calculation shows not only that $\Gamma \cdot \mathcal{K} = \mathcal{K}$, but also that this action extends the action of Γ on $\mathcal{Y}$.

Exercise 11.6.9 Show by example that the action (11.6.9) of Γ on its Busemann boundary $\partial_B \Gamma$ need not be transitive, even though the action on Γ itself *is* transitive.

HINT: Look again at Exercise 11.6.6.

Exercise 11.6.10 The *boundary cocycle* for the action of a group Γ on its Busemann boundary is the function $B : \Gamma \times \partial_B \Gamma \to \mathbb{Z}$ defined by

$$B(g, h) := -h(g^{-1}) \quad \text{for} \;\; g \in \Gamma \text{ and } h \in \mathcal{K}. \tag{11.6.11}$$

Check that this function satisfies the *cocycle identity*

$$B(g_1 g_2, h) = B(g_2, h) + B(g_1, g_2 \cdot h). \tag{11.6.12}$$

11.7 The Karlsson–Ledrappier Theorem

If $(X_n)_{n \geq 0}$ is a transient random walk on a finitely generated group Γ then the image trajectory $(\Phi_{X_n})_{n \geq 0}$ in the Busemann compactification $\overline{\Gamma}$ must accumulate at points of the boundary $\partial_B \Gamma$. It need not converge to a single point, however:

Exercise 11.7.1 Let $(X_n)_{n \geq 0}$ be simple random walk on $\mathbb{Z}^3$. Show that with probability one, *every* point of the Busemann compactification $\partial_B \mathbb{Z}^3$ is a limit point of the sequence $(\Phi_{X_n})_{n \geq 0}$.

HINT: There are two horofunctions for every line in $\mathbb{Z}^3$ parallel to one of the coordinate axes (see Exercise 11.6.6), and these are dense in the Busemann boundary. Thus, it suffices to show that with probability one, the random walk $(X_n)_{n \geq 0}$ visits each of these lines infinitely often at points in each of the two directions.

Nevertheless, for many interesting groups it will be the case that random walk paths converge almost surely in $\overline{\Gamma}$. Among such groups are the free groups $\mathbb{F}_k$, where $k \geq 2$, and (as we will show in Chapter 13) the *hyperbolic groups*.

Lemma 11.7.2 *Let* $\Phi : \Gamma \to \overline{\Gamma}$ *be the standard embedding* (11.6.1) *of the group* Γ *in its Busemann compactification* $\overline{\Gamma}$*, and let* $(X_n)_{n\geq 0}$ *be an irreducible random walk in* Γ *with step distribution* μ*. If the sequence* $(\Phi_{X_n})_{n\geq 0}$ *converges almost surely to a point* $Z \in \partial_B\Gamma$ *then the distribution* λ *of* Z *is* μ*-stationary and the pair* $(\partial_B\Gamma, \lambda)$ *is a* μ*-boundary.*

Proof. If the sequence $(\Phi_{X_n})_{n\geq 0}$ converges almost surely then the limit Z is a measurable function $Z = \zeta(\xi_1, x_2, \cdots)$ of the random walk increments. Because the embedding Φ satisfies the intertwining identity (11.6.10), the function ζ must satisfy the identity

$$\zeta(\xi_1, \xi_2, \cdots) = \xi_1 \cdot \zeta(\xi_2, \xi_3, \cdots);$$

hence, by Proposition 11.4.9, the pair $(\partial_B\Gamma, \lambda)$ is a μ-boundary. $\square$

Theorem 11.7.3 (Karlsson-Ledrappier [72, 74]) *Let* μ *be the step distribution of a transient, irreducible random walk* $(X_n)_{n\geq 0}$ *on a finitely generated group* Γ*, and assume that* μ *has finite first moment* $\sum_{g\in\Gamma} |g|\mu(G)$*. Assume further that there is a* μ*-stationary Borel probability measure* λ *on the Busemann boundary* $\partial_B\Gamma$ *such that the pair* $(\partial_B\Gamma, \lambda)$ *is a* μ*-boundary with boundary map* $\zeta : \Gamma^\infty \to \partial_B\Gamma$*. Then*

(A) the speed ℓ *of the random walk is*

$$\ell = \sum_{g\in\Gamma} \mu(g) \int_{\partial_B\Gamma} h(g^{-1})\, d\lambda(h); \quad \text{and} \tag{11.7.1}$$

(B) with probability one,

$$\lim_{n\to\infty} \frac{-Z(X_n)}{n} = \ell \tag{11.7.2}$$

where Z *is the (random) horofunction* $Z = \zeta(\xi_1, \xi_2, \cdots)$ *and* $\xi_1, \xi_2, \cdots$ *are the increments of the random walk.*

The assumption that the step distribution has finite first moment ensures that the random walk has finite speed — see Corollary 3.3.2. For any horofunction h and any $g \in \Gamma$ the inequality $-h(g) \leq |g|$ holds, by (11.6.4). Thus, the content of the assertion (11.7.2) is that the boundary map ζ isolates a *particular* horofunction Z along which approximate equality $-Z(X_n) \sim |X_n|$ holds.

Example 11.7.4 By Exercise 11.6.5, the Busemann boundary of the free group $\mathbb{F}_k$ on $k \geq 2$ generators is homeomorphic to the space $\partial\mathbb{F}_k$ of ends of the Cayley graph. Thus, for any irreducible random walk $(X_n)_{n\geq 0}$ in $\mathbb{F}_k$ the limit $Z = \lim_{n\to\infty} \Phi_{X_n}$ exists almost surely, and so Lemma 11.7.2 applies. In this case Theorem 11.7.3

reduces to Furstenberg's Theorem 11.5.2. Oddly enough, on the (almost sure) event that $Z = \lim_{n\to\infty} \Phi_{X_n}$ exists, for any horofunction $h \neq Z$,

$$\lim_{n\to\infty} \frac{h(X_n)}{n} = \ell. \tag{11.7.3}$$

(Note the sign change from (11.7.2)!)

Even without the hypothesis that there is a μ-stationary measure λ on $\partial_B\Gamma$ such that $(\partial_B\Gamma, \lambda)$ is a μ-boundary, there is an analogue of the identity (11.7.1).

Theorem 11.7.5 (Karlsson-Ledrappier [72, 74]) *For any irreducible, transient random walk $(X_n)_{n\geq 0}$ on a finitely generated group Γ whose step distribution μ has finite first moment there is a μ-stationary Borel probability measure λ on the Busemann boundary $\partial_B\Gamma$ such that the identity* (11.7.1) *holds.*

Proof of Theorem 11.7.5. Let $\Phi : \Gamma \to \overline{\Gamma}$ be the standard embedding (11.6.1) of Γ in its Busemann compactification, and for each $n \in \mathbb{N}$ set

$$\lambda_n = \frac{1}{n}\sum_{k=0}^{n-1} \nu_k$$

where ν_k is the distribution of Φ_{X_k}. The compactness of $\overline{\Gamma}$, together with the Helly Selection Principle, implies that the sequence $(\lambda_n)_{n\geq 1}$ has a weakly convergent subsequence $(\lambda_{n_m})_{m\in\mathbb{N}}$ with limit λ. Because the random walk $(X_n)_{n\geq 0}$ is transient,the measure λ must have support contained in $\partial_B\Gamma$. We will show that λ satisfies equation (11.7.1).

By hypothesis, the increments ξ_i of the random walk are independent and identically distributed, so for any $0 \leq k < n$ the probability measure ν_k is the distribution of $\Phi_{X_{n-k}^{-1}X_n}$. Now for any $g \in \Gamma$ and any $k \in \mathbb{N}$,

$$\begin{aligned}
-\Phi_{X_{n-k}^{-1}X_n}(g^{-1}) &= d(X_{n-k}^{-1}X_n, g^{-1}) - d(X_{n-k}^{-1}X_n, 1) \\
&= d(gX_{n-k}^{-1}X_n, 1) - d(X_{n-k}^{-1}X_n, 1) \\
&= |g\xi_{n-k+1}\xi_{n-k+2}\cdots\xi_n| - |\xi_{n-k+1}\xi_{n-k+2}\cdots\xi_n|.
\end{aligned}$$

Applying this with $g = \xi_{n-k}$, we obtain

$$\sum_{g\in\Gamma}\mu(g)\int_{\overline{\Gamma}} h(g^{-1})\,d\nu_k(h) = E(|X_{n-k-1}^{-1}X_n| - E|X_{n-k}^{-1}X_n|,$$

which implies that for every $n \in \mathbb{N}$,

$$\sum_{g \in \Gamma} \mu(g) \int_{\overline{\Gamma}} h(g^{-1}) \, d\lambda_n(h) = \frac{E|X_n|}{n}.$$

But for any $g \in \Gamma$ the function $F_g : \overline{\Gamma} \to \mathbb{R}$ that maps $h \mapsto h(g^{-1})$ is continuous, so by definition of weak convergence,

$$\begin{aligned}
\ell &= \lim_{m \to \infty} \frac{E|X_{n_m}|}{n_m} \\
&= \lim_{m \to \infty} \sum_{g \in \Gamma} \mu(g) \int_{\overline{\Gamma}} h(g^{-1}) \, d\lambda_{n_m}(h) \\
&= \sum_{g \in \Gamma} \mu(g) \int_{\overline{\Gamma}} h(g^{-1}) \, d\lambda(h).
\end{aligned}$$

□

Proof of Theorem 11.7.3. Transitivity of the Γ-action on $\partial_B \Gamma$ ensures that the μ-stationary probability measure λ is unique, by Proposition 11.4.13. Therefore, Theorem 11.7.5 implies the equality (11.7.1).

For each nonnegative integer n set $Z_n := \zeta(\xi_{n+1}, \xi_{n+2}, \cdots)$, where ζ is the boundary map and $\xi_1, \xi_2, \cdots$ are the increments of the random walk. By Proposition 11.4.7, the sequence $(\xi_n, Z_n)_{n \geq 1}$ is stationary and ergodic. Let B : $\Gamma \times \partial_B \Gamma$ be the boundary cocycle defined by equation (11.6.11); then by the cocycle identity (11.6.12), for any $n \in \mathbb{N}$,

$$\begin{aligned}
-Z(X_n) &= B(\xi_n^{-1} \xi_{n-1}^{-1} \cdots \xi_1^{-1}, Z) \\
&= \sum_{i=1}^{n} B(\xi_{i+1}^{-1}, \xi_i^{-1} \xi_{i-1}^{-1} \cdots \xi_1^{-1} \cdot Z) \\
&= \sum_{i=1}^{n} B(\xi_i^{-1}, Z_i).
\end{aligned}$$

Therefore, the Ergodic Theorem 2.1.6 implies that

$$\lim_{n \to \infty} \frac{-Z(X_n)}{n} = EB(\xi_1^{-1}, Z_1) = \sum_{g \in \Gamma} \mu(g) \int_{\partial_B \Gamma} h(g^{-1}) \, d\lambda(h) \quad \text{almost surely.}$$

□

If a symmetric random walk with finitely supported step distribution has the Liouville property, then it must have speed $\ell = 0$, by Proposition 4.1.3 and Theorem 10.1.1. However, a random walk with asymmetric step distribution can

have positive speed and still satisfy the Liouville property. For such random walks, the Karlsson-Ledrappier theorems have the following interesting consequence.

Corollary 11.7.6 (Karlsson and Ledrappier [73]) *Let $(X_n)_{n\geq 0}$ be a transient, irreducible random walk on a finitely generated group Γ whose step distribution μ has finite support. If there are no bounded, nonconstant μ-harmonic functions on Γ then there exists a 1-Lipschitz homomorphism $\varphi : \Gamma \to \mathbb{R}$ such that with probability one,*

$$\lim_{n\to\infty} \frac{\varphi(X_n)}{n} = \ell := \lim_{n\to\infty} \frac{|X_n|}{n}. \tag{11.7.4}$$

Proof. Let λ be a μ-stationary Borel probability measure on $\partial_B \Gamma$ for which the equality (11.7.1) holds. The existence of such a measure is guaranteed by Theorem 11.7.5. Since there are no nonconstant bounded harmonic functions, Proposition 11.2.5 and the Riesz-Markov Theorem imply that the measures $g \cdot \lambda$, where $g \in \Gamma$, are all identical, equivalently, the measure λ is Γ-invariant.

Define

$$\varphi(g) = \int_{\partial_B \Gamma} h(g^{-1})\, d\lambda(h) = -\int_{\partial_B \Gamma} B(g,h)\, d\lambda(h), \tag{11.7.5}$$

where $B(\cdot,\cdot)$ is the boundary cocycle (11.6.11). Since $|B(g,h)| = |h(g)| \leq |g|$, the function φ satisfies $|\varphi(g)| \leq |g|$ for all $g \in \Gamma$.

By the cocycle identity (11.6.12), for any $g_1, g_2 \in \Gamma$,

$$\begin{aligned}
\varphi(g_1 g_2) &= -\int_{\partial_B \Gamma} B(g_2,h)\, d\lambda(h) - \int_{\partial_B \Gamma} B(g_1, g_2 \cdot h)\, d\lambda(h) \\
&= -\int_{\partial_B \Gamma} B(g_2,h)\, d\lambda(h) - \int_{\partial_B \Gamma} B(g_1,h)\, d(g_2 \cdot \lambda)(h) \\
&= -\int_{\partial_B \Gamma} B(g_2,h)\, d\lambda(h) - \int_{\partial_B \Gamma} B(g_1,h)\, d\lambda(h) \\
&= \varphi(g_1) + \varphi(g_2);
\end{aligned}$$

here we have used the Γ-invariance of λ in the third equality. Thus, φ is a group homomorphism, and since $|\varphi(g)| \leq |g|$ for all $g \in \Gamma$, it follows that φ is 1-Lipschitz relative to the word metric.

Finally, because φ is a homomorphism, the sequence $(\varphi(X_n))_{n\geq 0}$ is a random walk on $\mathbb{R}$, so the Strong Law of Large Numbers implies that with probability one,

$$\lim_{n\to\infty} \frac{\varphi(X_n)}{n} = E\varphi(X_1) = \sum_{g\in\Gamma} \mu(g) \int_{\partial_B \Gamma} h(g^{-1})\, d\lambda(h) = \ell.$$

□

Additional Notes. The notion of a μ-boundary originates in the seminal article [47] of Furstenberg, as do most of the results of Section 11.4. Proposition 11.3.4 and Theorem 11.5.2 are also taken from this paper. The speed of a random walk with finitely supported step distribution on a free product group can be evaluated explicitly: see Mairesse and Mathéus [95]. Theorem 11.5.4 is a special case of a general theorem of Furstenberg [45] for random products in a semi-simple Lie group. Bougerol [17], Chapter II, gives a lucid account of the natural generalization to random products in $SL(2, \mathbb{R})$. Theorems 11.7.3 and 11.7.5 are due to Karlsson and Ledrappier, as is Corollary 11.7.6: see their papers [72], [74], and [73]. The arguments of these articles also lead to a "geometric" proof of Osseledec's Multiplicative Ergodic Theorem. This is nicely explained in the recent survey article of Filip [40]. For an extensive (but highly non-elementary) treatment of the limit theory of random walks on matrix groups, see Benoist & Quint [10].

Chapter 12
Poisson Boundaries

A group whose Cayley graphs is a tree, such as the free group $\mathbb{F}_d$ on d generators, has a natural *geometric* boundary, the space of *ends* of the tree. Every random walk trajectory converges to a unique end (cf. Exercises 1.6.5 and 1.6.7), and bounded harmonic functions are in one-to-one correspondence with bounded functions on the space of ends (cf. Exercise 9.3.2). Is there an analogue of this geometric boundary for other finitely generated groups? H. Furstenberg [47], [48], [46] showed that — at least in an abstract sense — there is, and also that for *matrix groups* the boundary can be described in an explicit fashion. This chapter is an introduction to Furstenberg's theory.

Assumption 12.0.1 *Assume throughout this chapter that* $\mathbf{X} = (X_n)_{n\geq 0}$ *is a transient, irreducible random walk on a finitely generated group* Γ *with step distribution* μ *and increments* $(\xi_n)_{n\geq 1}$. *Let* $\mathcal{I}$ *be the* σ*-algebra of invariant events in* Γ^∞ *and* ν_x *the exit measure for the initial state* x *(cf. Definition 9.1.3), with the abbreviation* $\nu = \nu_1$.

12.1 Poisson and Furstenberg-Poisson Boundaries

In Chapter 9, we saw that for any irreducible random walk on a finitely generated group Γ the space of bounded harmonic functions is linearly isomorphic to the Banach space $L^\infty(\Gamma^\infty, \mathcal{I}, \nu)$ of bounded, $\mathcal{I}$-measurable random variables (cf. Theorem 9.1.5 and Corollary 9.2.3). We say that the probability space $(\Gamma^\infty, \mathcal{I}, \nu)$ is a *Poisson boundary* for the random walk $(X_n)_{n\geq 0}$; see Definition 12.1.11 below.

The probability space $(\Gamma^\infty, \mathcal{I}, \nu)$ is a purely measure-theoretic object; there is no topology attached. Consequently, there can be no sense in which random walk paths "converge to points of the boundary", as there is for random walks in free groups or free products. Furstenberg's original conception of a Poisson boundary required a topological as well as a measure space structure, with the topology subject to the

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_12

requirement that random walk paths should converge almost surely. For this reason, we make the following definition.

Definition 12.1.1 A *Furstenberg-Poisson boundary* for a random walk $(X_n)_{n\geq 0}$ with step distribution μ is a μ-boundary $(\mathcal{Y}, \lambda)$ such that

(i) the group Γ acts transitively on $\mathcal{Y}$; and
(ii) every bounded, μ-harmonic function $u : \Gamma \to \mathbb{R}$ has a representation

$$u(g) = \int_{\mathcal{Y}} f(g \cdot y)\, d\lambda(y) = \int_{\mathcal{Y}} f(y)\, d(g \cdot \lambda)(y) \quad \forall\, g \in \Gamma \tag{12.1.1}$$

for some bounded, Borel measurable function $f : \mathcal{Y} \to \mathbb{R}$.

Here $g \cdot \lambda$ refers to the induced Γ-action on the space of Borel probabilities: see Definition 11.2.1. Recall (cf. Proposition 11.2.5) that if $\mathcal{Y}$ is a Γ-space then for any bounded, measurable function $f : \mathcal{Y} \to \mathbb{R}$ the function $u : \Gamma \to \mathbb{R}$ defined by equation (12.1.1) is harmonic. Later in this section we will prove that the integral representation (12.1.1) is (essentially) unique.

What justifies calling an object $(\mathcal{Y}, \lambda)$ that satisfies the requirements of Definition 12.1.1 a *boundary*? For this, we need a topology, specifically, a topological embedding of the group Γ, with its discrete topology, in a compact space $\Gamma \cup \mathcal{Z}$, with $\mathcal{Y}$ embedded in $\mathcal{Z}$. If $(\mathcal{Y}, \lambda)$ is a μ-boundary, as we shall henceforth assume, then a natural approach is to identify points of $\Gamma \cup \mathcal{Y}$ with Borel probability measures on $\mathcal{Y}$ by the rule

$$\begin{aligned} g &\mapsto g \cdot \lambda \quad \text{if } \ g \in \Gamma, \\ y &\mapsto \delta_y \quad \text{if } \ y \in \mathcal{Y}, \end{aligned} \tag{12.1.2}$$

and then to restrict the topology of weak convergence to the closure of the set $\{g \cdot \lambda\}_{g \in \Gamma}$. Unfortunately, the mapping from Γ into the space $\mathcal{M}(\mathcal{Y})$ of Borel probability measures on $\mathcal{Y}$ defined by $g \mapsto (g \cdot \lambda)$ need not be injective.

To circumvent this difficulty, we will artificially enhance the weak topology on measures so as isolate the points of Γ, as follows. Let ρ_w be a metric for the topology of weak convergence on $\mathcal{M}(\mathcal{Y})$; that such a metric exists follows from the separability of $C(\mathcal{Y})$ (see Corollaries A.7.6 and A.7.8 of the Appendix). Let $\Gamma = \{g_n\}_{n \in \mathbb{N}}$ be an enumeration of Γ, and define a distance ρ on $\Gamma \cup \mathcal{M}(\mathcal{Y})$ as follows:

$$\begin{aligned} \rho(\nu, \nu') &= \rho_w(\nu, \nu') && \text{for } \nu, \nu' \in \mathcal{M}(\mathcal{Y}); \\ \rho(\nu, g_n) &= \rho_w(\nu, g_n \cdot \lambda) + 2^{-n} && \text{for } \ g_n \in \Gamma, \nu \in \mathcal{M}(\mathcal{Y}); \\ \rho(g_n, g_m) &= \rho_w(g_n \cdot \lambda, g_m \cdot \lambda) + 2^{-n} + 2^{-m} && \text{for } \ g_n \neq g_m \in \Gamma. \end{aligned} \tag{12.1.3}$$

Exercise 12.1.2 Verify that ρ is a metric on $\Gamma \cup \mathcal{M}(\mathcal{Y})$ whose restriction to Γ induces the discrete topology, and show that the metric space $(\Gamma \cup \mathcal{M}(\mathcal{Y}), \rho)$ is compact.

HINT: See Theorem A.7.9 of the Appendix.

Proposition 12.1.3 *If $(\mathcal{Y}, \lambda)$ is a μ-boundary then for every $x \in \Gamma$*

$$P^x \left\{ \lim_{n\to\infty} \rho(X_n, \delta_Z) = 0 \right\} = 1 \tag{12.1.4}$$

where δ_Z is the almost sure weak limit of the sequence $(X_n \cdot \lambda)_{n\geq 0}$.

Proof. This is merely a restatement of Proposition 11.4.9, which asserts that with probability one the sequence of Borel probability measures $X_n \cdot \lambda$ converge weakly to the point mass δ_Z. Since the random walk $(X_n)_{n\geq 0}$ is transient, weak convergence of the sequence $(X_n \cdot \lambda)_{n\geq 0}$ is equivalent to convergence in the metric ρ. □

Theorem 12.1.4 (Furstenberg [47]) *Every irreducible random walk on an infinite, finitely generated group has a Furstenberg-Poisson boundary.*

The proof is deferred to Section 12.4. The existence of Furstenberg-Poisson boundary has some important theoretical consequences: see, for instance, Proposition 12.4.4 in Section 12.4 below. In most circumstances, however, it isn't the mere *existence* of a Furstenberg-Poisson boundary that is important, but rather the *identification* of such a boundary as a recognizable metric space with a natural Γ-action. It is often easy to find Γ-spaces — that is, compact metric spaces on which the group Γ acts by homeomorphisms — and Proposition 11.2.4 ensures the existence of μ-stationary Borel probability measures on such spaces. Lemma 11.4.8 and Proposition 11.4.9 provide conditions for ascertaining if a Γ-space $\mathcal{Y}$ admits a stationary probability measure λ such that the pair $(\mathcal{Y}, \lambda)$ is a μ-boundary. The remaining issue is that of determining when a μ-boundary $(\mathcal{Y}, \lambda)$ is a Furstenberg-Poisson boundary.

This problem, unlike that of random walk path convergence to the boundary, is of a purely measure-theoretic nature: it is equivalent (as we will see) to finding conditions under which

$$L^\infty(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}}, \lambda) \cong L^\infty(\Gamma^\infty, \mathcal{I}, \nu). \tag{12.1.5}$$

The key is that for every μ-boundary there is a function $Y : \Gamma^\infty \to \mathcal{Y}$, the associated *equivariant boundary map*, that satisfies

(EQ-1) Y is $\mathcal{I}$-measurable;
(EQ-2) the distribution of $Z := Y(X_0, X_1, \cdots)$ under P is λ; and
(EQ-3) for every $g \in \Gamma$, with P-probability one,

$$Y(g, gX_1, gX_2, \cdots) = g \cdot Y(1, X_1, X_2, \cdots). \tag{12.1.6}$$

Definition 12.1.5 A probability space $(\Upsilon, \mathcal{G}, \lambda)$ is a *measurable μ-boundary* if

(A) Γ acts by measurable transformations on $(\Upsilon, \mathcal{G})$; and
(B) there exists an $\mathcal{I}$-measurable mapping $Y : \Gamma^{\infty} \to \Upsilon$, called the associated *equivariant boundary map*, satisfying the conditions (EQ-1), (EQ-2), and (EQ-3) above.

Example 12.1.6 Any μ-boundary (in the sense of Definition 11.4.6) is a measurable μ-boundary, because an action of Γ on a metric space $\mathcal{Y}$ by homeomorphisms is also an action by (Borel) measurable transformations.

If $Y : \Gamma^{\infty} \to \Upsilon$ is shift-invariant (that is, if $Y = Y \circ \sigma$) then it is $\mathcal{I}$-measurable, but the converse need not be true. The following example points to an important instance of a measurable μ-boundary whose associated equivariant boundary map Y is not shift-invariant.

Example 12.1.7 Let $\mathcal{G}$ be a σ-algebra contained in the invariant σ-algebra $\mathcal{I}$ on Γ^{∞}, and let $Y : \Gamma^{\infty} \to \Gamma^{\infty}$ be the identity map. Since $\mathcal{G} \subset \mathcal{I}$, the identity mapping $Y : (\Gamma^{\infty}, \mathcal{I}) \to (\Gamma^{\infty}, \mathcal{G})$ is a measurable transformation, and clearly the distribution of $Y(X_0, X_1, \cdots)$ under P is the restriction of the exit measure ν to the σ-algebra $\mathcal{G}$. Thus, both (EQ-1) and (EQ-2) hold. The group Γ acts on the sequence space Γ^{∞} by left multiplication: in particular, if $\mathbf{x} = (x_0, x_1, x_2, \cdots)$ and $g \in \Gamma$, then

$$g \cdot \mathbf{x} := (gx_0, gx_1, gx_3, \cdots). \tag{12.1.7}$$

This action commutes with the identity map Y, so (EQ-3)holds; therefore, the probability space $(\Gamma^{\infty}, \mathcal{G}, \nu)$ is a measurable μ-boundary with associated equivariant boundary map Y.

As with (topological) μ-boundaries, the Γ-action on a measurable μ-boundary $(\Upsilon, \mathcal{G}, \lambda)$ induces an action on the space of probability measures on $\mathcal{G}$. The action is defined precisely as in Definition 11.2.1: if $g \in \Gamma$ and α is a probability measure on $\mathcal{G}$, then for any bounded, $\mathcal{G}$-measurable function $f : \Upsilon \to \mathbb{R}$,

$$\int_{\Upsilon} f(y)\, d(g \cdot \alpha)(y) := \int_{\Upsilon} f(g \cdot y)\, d\alpha(y). \tag{12.1.8}$$

We will denote the probability measures in the Γ-orbit of λ by subscripts: thus,

$$\lambda_g := g \cdot \lambda.$$

Proposition 12.1.8 *Let $(\Upsilon, \mathcal{G}, \lambda)$ be a measurable μ-boundary and $Y : \Gamma^{\infty} \to \Upsilon$ the associated equivariant boundary map. Then for every $g \in \Gamma$,*

$$\lambda_g = \nu_g \circ Y^{-1} \tag{12.1.9}$$

where $\lambda_g := g \cdot \lambda$ and $\nu_g = g \cdot \nu$ is the exit measure for the random walk with initial point g (that is, ν_g is the restriction of $\hat{P}^g$ to the invariant σ-algebra $\mathcal{I}$, where $\hat{P}^g$ is the law of the random walk with initial point g).

Proof. Consider first the special case $g = 1$. Since $Z = Y \circ \mathbf{X}$ has distribution λ under $P = P^1$, for any event $B \in \mathcal{G}$,

$$\begin{aligned}\lambda(B) &= P\{Y \circ \mathbf{X} \in B\} \\ &= P\left\{\mathbf{X} \in Y^{-1}(B)\right\} \\ &= \hat{P}(Y^{-1}(B)) \\ &= \nu(Y^{-1}(B)).\end{aligned}$$

The final equality holds because the event $Y^{-1}(B)$ is an element of the σ-algebra $\mathcal{I}$. Thus, the identity (12.1.9) holds for the group identity $g = 1$.

The general case follows by left translation, using the fact that the map Y is equivariant. Let $f : \Upsilon \to \mathbb{R}$ be any bounded, measurable function, and for any $g \in \Gamma$ define f_g to be the translated function $f_g(y) = f(g \cdot y)$. Then

$$\int f(y)\, d\lambda_g(y) = \int f(g \cdot y)\, d\lambda(y) = \int f_g(y)\, d\lambda(y).$$

But $\lambda = \nu \circ Y^{-1}$, so

$$\begin{aligned}\int f_g(y)\, d\lambda(y) &= \int f_g(Y(\mathbf{x}))\, d\nu(\mathbf{x}) \\ &= \int f(g \cdot Y(\mathbf{x}))\, d\nu(\mathbf{x}) \\ &= \int f(Y(g \cdot \mathbf{x}))\, d\nu(\mathbf{x}) \quad \text{(by (12.1.6))} \\ &= \int f(Y(\mathbf{x}))\, d\nu_g(\mathbf{x}).\end{aligned}$$

Here $g \cdot \mathbf{x}$ denotes the action (12.1.7) of Γ on Γ^∞ by left multiplication: in particular, if $\mathbf{x} = (x_0, x_1, \cdots)$ then $g \cdot \mathbf{x} = (gx_0, gx_1, \cdots)$. This shows that for every bounded, measurable function f,

$$\int f(y)\, d\lambda_g(y) = \int f(Y(\mathbf{x}))\, d\nu_g(\mathbf{x}),$$

and so (12.1.9) follows. $\square$

Corollary 12.1.9 *Under the hypotheses of Proposition 12.1.8, the measures λ_g are mutually absolutely continuous, and their likelihood ratios satisfy the identity*

$$\frac{d\lambda_g}{d\lambda} \circ Y = \left(\frac{d\nu_g}{d\nu}\right)_{\sigma(Y)} \qquad \nu - \text{almost surely}, \tag{12.1.10}$$

where $\sigma(Y) \subset \mathcal{I}$ is the σ-algebra generated by the random variable Y.

Proof. Absolute continuity follows from (12.1.9), because for an irreducible random walk the exit measures are mutually absolute continuous, by Proposition 9.2.2. The formula (12.1.10) is a routine exercise in measure theory (see Exercise A.8.8 of the Appendix), using (12.1.9). □

Corollary 12.1.10 *Under the hypotheses of Proposition 12.1.8, the mapping $f \mapsto f \circ Y$ is a linear isomorphism of the Banach space $L^\infty(\Upsilon, \mathcal{G}, \lambda)$ onto $L^\infty(\Gamma^\infty, \sigma(Y), \nu)$. Consequently, for any bounded, μ-harmonic function $u : \Gamma \to \mathbb{R}$ with a representation (12.1.1), the representation is (essentially) unique.*

Proof. The mapping $f \mapsto f \circ Y$ is linear and one-to-one. To prove that it is surjective, it suffices (by linearity and the usual device of approximation by simple functions) to show that for every event $F \in \sigma(Y)$ there is an event $B \in \mathcal{G}$ such that $\mathbf{1}_F = \mathbf{1}_B \circ Y$. But this is a direct consequence of the definition of the σ-algebra $\sigma(Y)$ as the minimal σ-algebra containing all event of the form $\{Y \in B\}$.

Uniqueness of integral representations (12.1.1) now follows by the uniqueness assertion in Blackwell's Theorem 9.1.5, as the measures ν_x are all mutually absolutely continuous. □

Definition 12.1.11 A *Poisson boundary* of the random walk $(X_n)_{n\geq 0}$ (or its law) is a measurable μ-boundary $(\Upsilon, \mathcal{G}, \lambda)$ such that

$$L^\infty(\Upsilon, \mathcal{G}, \lambda) \cong L^\infty(\Gamma^\infty, \mathcal{I}, \nu). \tag{12.1.11}$$

When is a measurable μ-boundary a Poisson boundary? We have already addressed this issue, in Section 9.3, where we found that a necessary and sufficient condition for equality in (12.1.11) is that the likelihood ratios $L_x := (d\nu_x/d\nu)$ all have $\mathcal{G}$-measurable versions. This, in conjunction with Corollary 12.1.9, gives us the following criterion.

Corollary 12.1.12 *A measurable μ-boundary $(\Upsilon, \mathcal{G}, \lambda)$ with associated equivariant boundary map $Y : \Gamma^\infty \to \Upsilon$ is a Poisson boundary if and only if*

$$L_g = \frac{d\nu_g}{d\nu} = \frac{d\lambda_g}{d\lambda} \circ Y \quad \nu - \text{almost surely}. \tag{12.1.12}$$

□

Exercise 12.1.13 Let $(\Upsilon, \mathcal{G}, \lambda)$ be a measurable μ-boundary, and for each $g \in \Gamma$ write $\lambda_g = (g \cdot \lambda)$. Define

$$\varphi(g, y) := \frac{d\lambda_{g^{-1}}}{d\lambda}(y) \tag{12.1.13}$$

to be the likelihood ratio of λ_g with respect to $\lambda = \lambda_1$ at the point $y \in \Upsilon$. Show that φ is a *multiplicative cocycle*, that is, for any $g_1, g_2 \in \Gamma$ and $y \in \Upsilon$,

$$\varphi(g_1 g_2, y) = \varphi(g_1, g_2 \cdot y)\varphi(g_2, y). \tag{12.1.14}$$

HINT: See Exercise A.8.7 of the Appendix.

NOTE: Since likelihood ratios are only defined up to sets of measure 0, the definition of the cocycle φ is ambiguous; however, since for each pair g_1, g_2 the equality (12.1.14) fails only on a set of measure zero, and since there are only countably many pairs g_1, g_2, versions of the likelihood ratios (12.1.13) can be chosen so that (12.1.14) holds everywhere.

12.2 Entropy of a Boundary

Direct verification of the criterion (12.1.12) is not always feasible. There is, however, an alternative criterion based on an entropy invariant. Recall that for two mutually absolutely continuous probability measures λ_1, λ_2 on a measurable space $(\Upsilon, \mathcal{G})$, the *Kullback-Leibler divergence* $D(\lambda_2 \parallel \lambda_1)$ is the nonnegative real number

$$D(\lambda_2 \parallel \lambda_1) := -\int \log \left(\frac{d\lambda_1}{d\lambda_2}\right)_{\mathcal{G}} d\lambda_2. \tag{12.2.1}$$

If necessary, the dependence on the σ-algebra $\mathcal{G}$ will be indicated by affixing a subscript:

$$D(\lambda_2 \parallel \lambda_1)_{\mathcal{G}} := D(\lambda_2 \parallel \lambda_1). \tag{12.2.2}$$

Definition 12.2.1 The *Furstenberg entropy* $h_F(\mu; \lambda)$ of a measurable μ-boundary $(\Upsilon, \mathcal{G}, \lambda)$ is defined by

$$h_{\mathrm{F}}(\mu; \lambda) := \sum_{g \in \Gamma} \mu(g) D(\lambda_g \parallel \lambda) = \sum_{g \in \Gamma} \mu(g) D(\lambda \parallel \lambda_{g^{-1}}), \tag{12.2.3}$$

or equivalently

$$h_{\mathrm{F}}(\mu; \lambda) = -E \log \left(\frac{d\lambda_{X_1}}{d\lambda}\right)(Y \circ \mathbf{X}) \tag{12.2.4}$$

where $Y : \Gamma^\infty \to \Upsilon$ is the associated equivariant boundary map.

We will show in Theorem 12.2.4 below that if the random walk $(X_n)_{n\geq 0}$ has finite Avez entropy then any measurable μ-boundary also has finite Furstenberg entropy. If $(\Upsilon, \mathcal{G}, \lambda)$ is a measurable μ-boundary, then it is also a measurable μ^{*m}-boundary for any positive integer m, because if λ is μ-stationary then it is evidently also μ^{*m}-stationary.

Lemma 12.2.2 *Under the hypotheses of Theorem 12.2.3, for any integer $m \geq 1$,*

$$h_F(\mu^{*m}; \lambda) = m h_F(\mu; \lambda). \tag{12.2.5}$$

Proof. If $(X_n)_{n\geq 0}$ is a random walk with step distribution μ then $(X_{mn})_{n\geq 0}$ is a random walk with step distribution μ^{*m}. Let $\xi_n = X_{n-1}^{-1} X_n$ be the increments, and let $(\Omega, \mathcal{F}, P)$ be the probability space on which the random walk is defined. Define random variables $Z_n : \Omega \to \Upsilon$ by

$$Z_n = Y(1, \xi_{n+1}, \xi_{n+1}\xi_{n+2}, \cdots);$$

thus, $Z_0 = Y \circ \mathbf{X}$. The sequence $((\xi_n, Z_n))_{n\geq 0}$ is stationary (cf. Proposition 11.4.7 for the special case where $(\Upsilon, \mathcal{G}, \lambda)$ is a topological μ-boundary; the general case follows by the same argument), and the distribution of each term Z_n is λ. Consequently, by equation (12.1.6), for any $g \in \Gamma$ the random variable $g \cdot Z_0$ has distribution λ_g.

The measures $\lambda_g = g \cdot \lambda$ are mutually absolutely continuous, by Proposition 12.1.8, so

$$\frac{d\lambda_{X_n}}{d\lambda}(Z_0) = \prod_{m=1}^{n} \frac{d\lambda_{X_m}}{d\lambda_{X_{m-1}}}(Z_0).$$

By an elementary transformation law for Radon-Nikodym derivatives (cf. Exercise 12.1.13 (B), or, alternatively, Exercise A.8.7 in Section A.8 of the Appendix), the factors in the product can be re-expressed as

$$\frac{d\lambda_{X_m}}{d\lambda_{X_{m-1}}}(Z_0) = \frac{d\lambda_{\xi_m}}{d\lambda}(X_{m-1}^{-1} \cdot Z_0) = \frac{d\lambda_{\xi_m}}{d\lambda}(Z_{m-1}),$$

and hence,

$$\frac{d\lambda_{X_n}}{d\lambda}(Z_0)) = \prod_{m=1}^{n} \frac{d\lambda_{\xi_m}}{d\lambda}(Z_{m-1}). \tag{12.2.6}$$

It follows by stationarity of the sequence $((\xi_n, Z_n))_{n\geq 0}$ that

$$h_F(\mu^{*m}; \lambda) = -E \log\left(\frac{d\lambda_{X_m}}{d\lambda}\right)(Z_0)$$

$$
\begin{aligned}
&= -\sum_{n=1}^{m} E \log \left(\frac{d\lambda_{\xi_n}}{d\lambda} \right) (Z_{n-1}) \\
&= -mE \log \left(\frac{d\lambda_{\xi_1}}{d\lambda} \right) (Z_0) \\
&= m h_F(\mu; \lambda).
\end{aligned}
$$

□

Furstenberg showed (at least in a special case — see [47]) that entropy measures the exponential decay of likelihood ratios of the measures λ_x along random walk paths. Recall (cf. Assumption 12.0.1) that the random walk $\mathbf{X} = (X_n)_{n \geq 0}$ is irreducible and that its step distribution μ has finite Shannon entropy.

Theorem 12.2.3 (Furstenberg [47]) *If* $(\Upsilon, \mathcal{G}, \lambda)$ *is a measurable* μ*-boundary boundary with associated equivariant boundary map* $Y : \Gamma^\infty \to \Upsilon$ *and finite Furstenberg entropy* $h_F(\mu; \lambda)$*, then*

$$
\lim_{n \to \infty} n^{-1} \log \left(\frac{d\lambda_{X_n}}{d\lambda} \right) \circ Y \circ \mathbf{X} = h_F(\mu; \lambda) \quad P - \textit{almost surely.} \tag{12.2.7}
$$

Proof. As in the proof of Lemma 12.2.2, set

$$
Z_n = Y(1, \xi_{n+1}, \xi_{n+1}\xi_{n+2}, \cdots),
$$

where $\xi_1, \xi_2, \cdots$ are the increments of the random walk $(X_n)_{n \geq 0}$. Kolmogorov's 0–1 Law implies that the sequence $((\xi_n, Z_n))_{n \geq 0}$ is not only stationary, but also ergodic (cf. Proposition 2.1.3). Consequently, by Birkhoff's Ergodic Theorem,

$$
\begin{aligned}
\frac{1}{n} \log \left(\frac{d\lambda_{X_n}}{d\lambda} \right) \circ Y \circ \mathbf{X} &= \frac{1}{n} \sum_{n=1}^{m} \log \left(\frac{d\lambda_{\xi_n}}{d\lambda} \right) (Z_{n-1}) \\
&\longrightarrow E \log \left(\frac{d\lambda_{\xi_1}}{d\lambda} \right) (Z_0) \quad P - \text{almost surely} \\
&= -h_F(\mu; \lambda).
\end{aligned}
$$

□

Theorem 12.2.4 (Kaimanovich & Vershik [68]) *The Furstenberg entropy* $h_F(\mu; \lambda)$ *of a measurable* μ*-boundary* $(\Upsilon, \mathcal{G}, \lambda)$ *is bounded above by the Avez entropy* $h = h(\Gamma; \mu)$*, and the two entropies are equal if and only if* $(\Upsilon, \mathcal{G}, \lambda)$ *is a Poisson boundary.*

Proof. Let $(\Upsilon, \mathcal{G}, \lambda)$ be a measurable μ-boundary with associated equivariant boundary map Y, and let $\sigma(Y) \subset \mathcal{I}$ be the σ-algebra generated by Y. By

Corollary 12.1.12, $(\Upsilon, \mathcal{G}, \lambda)$ is a Poisson boundary if and only if $(\Gamma^\infty, \sigma(Y), \nu)$ is also a Poisson boundary. Furthermore, Corollary 12.1.9, together with the definition (12.2.3), implies that the boundaries $(\Upsilon, \mathcal{G}, \lambda)$ and $(\Gamma^\infty, \sigma(Y), \nu)$ have the same Furstenberg entropy. Therefore, to prove the theorem it suffices to consider measurable μ-boundaries of the form $(\Gamma^\infty, \mathcal{J}, \nu)$, where $\mathcal{J}$ is a Γ-invariant σ-algebra contained in $\mathcal{I}$ and where the equivariant boundary map Y is the identity.

Corollary 10.6.7 implies that for the natural Poisson boundary $(\Gamma^\infty, \mathcal{I}, \nu)$ the Furstenberg entropy coincides with the Avez entropy. Consequently, by Lemma 12.2.2, for any integer $m \geq 1$ the Avez entropy h satisfies

$$mh = mh_F(\mu; \nu) = \sum_z \mu^{*m}(z) D(\nu_z \parallel \nu). \tag{12.2.8}$$

Lemma 12.2.2, together with Definition 12.2.1, also implies that for any integer $m \geq 1$ the Furstenberg entropy of the measure boundary $(\Gamma^\infty, \mathcal{J}, \nu)$ satisfies

$$mh_F(\mu; \nu \restriction \mathcal{J}) = \sum_{g \in \Gamma} \mu^{*m}(g) D(\nu_g \restriction \mathcal{J} \parallel \nu \restriction \mathcal{J}). \tag{12.2.9}$$

(Here $\nu_g \restriction \mathcal{J}$ denotes the restriction of the exit measure ν_g to the σ-algebra $\mathcal{J}$.) Since $\mathcal{J} \subset \mathcal{I}$, it follows by Exercise 10.6.4 that for every $g \in \Gamma$,

$$D(\nu_g \restriction \mathcal{J} \parallel \nu \restriction \mathcal{J}) \leq D(\nu_g \parallel \nu)$$

with strict inequality unless the likelihood ratio $d\nu_g/d\nu$ is essentially $\mathcal{J}$-measurable. This, together with the identities (12.2.8)–(12.2.9), proves that

$$h_F(\mu; \lambda) \leq h,$$

and in addition shows that the inequality is strict unless $d\nu_g/d\nu$ is essentially $\mathcal{J}$-measurable for every $g \in \Gamma$ such that $\mu^{*m}(g) > 0$ for some integer $m \geq 1$. Since by our standing assumptions the step distribution μ is irreducible, for every $g \in \Gamma$ there is at least one integer $m \geq 1$ such that $\mu^{*m}(g) > 0$; consequently, $h_F(\mu; \lambda) = h$ if and only if $d\nu_g/d\nu$ is essentially $\mathcal{J}$-measurable for *every* $g \in \Gamma$. Corollary 12.1.12 implies that this is a necessary and sufficient condition for $(\Gamma^\infty, \mathcal{J}, \nu)$ to be a Poisson measure boundary. □

Denote by $\mathfrak{S}$ the set of all finite subsets F of Γ, and let $2^{\mathfrak{S}}$ be the σ-algebra consisting of all subsets of $\mathfrak{S}$. Since $\mathfrak{S}$ is countable, the σ-algebra $2^{\mathfrak{S}}$ is generated by the singleton subsets of $\mathfrak{S}$; hence, if $(\Upsilon, \mathcal{G})$ is a measurable space then a function $T : \Upsilon \to \mathfrak{S}$ is measurable if and only if for each $F \in \mathfrak{S}$ the set $T^{-1}(F)$ is an element of $\mathcal{G}$.

Corollary 12.2.5 (Kaimanovich's Localization Criterion) *Let $(\Upsilon, \mathcal{G}, \lambda)$ be a measurable μ-boundary whose equivariant boundary map Y is* shift-invariant*, that is, $Y = Y \circ \sigma$. Assume that the random walk has finite Avez entropy, and that the*

step distribution μ *has finite first moment vis-a-vis the word metric on* Γ*, that is,*

$$\alpha := \sum_{g \in \Gamma} |g| \mu(g)| < \infty. \tag{12.2.10}$$

Suppose that for each $\varepsilon > 0$ *there are measurable functions* $B_n : \Upsilon \to \mathfrak{S}$ *such that for every* $y \in \Upsilon$ *the cardinality* $|B_n(y)|$ *satisfies*

$$|B_n(y)| \le e^{n\varepsilon}, \tag{12.2.11}$$

and

$$\limsup_{n\to\infty} P\{X_n \in B_n(Y(\mathbf{X}))\} > 0. \tag{12.2.12}$$

Then $(\Upsilon, \mathcal{G}, \lambda)$ *is a Poisson boundary.*

This is a variant of Kaimanovich's *ray criterion*[69]. In typical applications, such as that considered in Sections 13.5, the sets B_n are balls of radius $e^{n\varepsilon}$ centered at random points of Γ. In rough terms, the localization criterion states that a measurable μ-boundary $(\Upsilon, \mathcal{G}, \lambda)$ will be a Poisson boundary if there is enough information about the trajectory of the random walk $(X_n)_{n\ge 0}$ in the random variable Y to pin down the location of X_n in a set of subexponential size with high (or at least positive) probability.

In proving Corollary 12.2.5, we may assume without loss of generality that the probability space on which the random walk is defined is the sequence space $(\Gamma^\infty, \mathcal{B}_\infty, P = \hat{P})$, and that the random variables X_n are the coordinate evaluation maps on Γ^∞. Thus, $Y = Y(\mathbf{X})$.

Lemma 12.2.6 *Under the hypotheses of Corollary 12.2.5, for any* $n \ge 1$ *and any* $g \in \gamma$*,*

$$E(\mathbf{1}_{\{X_n=g\}} \,|\, \sigma(Y)) = \mu^{*n}(g)\left(\frac{d\lambda_g}{d\lambda}\right)(Y) \tag{12.2.13}$$

Proof. The random variable on the right side of (12.2.13) is obviously measurable with respect to $\sigma(Y)$, so by the definition (A.9.1) of conditional expectation it suffices to show that for every event $G \in \mathcal{G}$,

$$E\mathbf{1}_{\{X_n=g\}}\mathbf{1}_G(Y) = \mu^{*n}(g) E\left(\mathbf{1}_G(Y)\frac{d\lambda_g}{d\lambda}(Y)\right).$$

Since the boundary map Y is invariant by the shift, the Markov property implies that

$$\begin{aligned} E\mathbf{1}_{\{X_n=g\}}\mathbf{1}_G(Y) &= E\mathbf{1}_{\{X_n=g\}}\mathbf{1}_G(Y(X_n, X_{n+1}, \cdots)) \\ &= E\mathbf{1}_{\{X_n=g\}}\mathbf{1}_G(Y(g, g\xi_{n+1}, g\xi_{n+1}\xi_{n+2}\cdots)) \end{aligned}$$

$$
\begin{aligned}
&= P\{X_n = g\}\, E\mathbf{1}_G(g \cdot Y(\mathbf{X})) \\
&= \mu^{*n}(g) E\mathbf{1}_G(g \cdot Y).
\end{aligned}
$$

On the other hand, since Y has distribution λ, for any $g \in \Gamma$ the random variable $g \cdot Y$ has distribution λ_g, so

$$
E\mathbf{1}_G(Y)\frac{d\lambda_g}{d\lambda}(Y) = \int_{\mathcal{Y}} \mathbf{1}_G \frac{d\lambda_g}{d\lambda}\, d\lambda = \int_{\mathcal{Y}} \mathbf{1}_G\, d\lambda_g = E\mathbf{1}_G(g \cdot Y).
$$

□

Proof of Corollary 12.2.5. Fix $\varepsilon > 0$, and let $B_n : \Upsilon \to \mathfrak{S}$ be functions satisfying the conditions (12.2.11) and (12.2.12). Define $F_n(Y)$ and G_n to be the sets

$$
\begin{aligned}
F_n(Y) &= \left\{ g \in \Gamma \,:\, |g| \leq 2n\alpha \ \text{ and } \ \log \frac{d\lambda_g}{d\lambda}(Y) \in [nh_F - n\varepsilon, nh_f + n\varepsilon] \right\}, \\
G_n &= \left\{ g \in \Gamma \,:\, \log \mu^{*n}(g) \in [nh - \varepsilon, nh + \varepsilon] \right\},
\end{aligned}
$$

where $h_F = h_F(\mu; \lambda)$ is the Furstenberg entropy of the boundary and h is the Avez entropy of the random walk. The set G_n is non-random, but $B_n(Y)$ and $F_n(Y)$ are $\mathcal{I}$-measurable, $\mathfrak{S}$-valued random variables. By Theorem 12.2.3 and Corollaries 3.3.2 and 3.3.3, the events $\{X_n \in F_n(Y)\}$ and $\{X_n \in G_n\}$ both have probabilities converging to 1 as $n \to \infty$, so

$$
\limsup_{n\to\infty} P(X_n \in B_n(Y) \cap F_n(Y) \cap G_n) > 0. \tag{12.2.14}
$$

For any $g \in \Gamma$ the event $\{g \in B_n(Y) \cap F_n(Y)\}$ is an element of the σ-algebra $\sigma(Y)$; consequently, by the definition of conditional expectation and Lemma 12.2.6, for all sufficiently large n,

$$
\begin{aligned}
P\{X_n \in B_n(Y) \cap F_n(Y) \cap G_n\} &= \sum_{g \in G_n} E\mathbf{1}_{\{X_n = g\}} \mathbf{1}_{B_n(Y)\cap F_n(Y)}(g) \\
&= \sum_{g \in G_n} EE(\mathbf{1}_{\{X_n = g\}} \,|\, \sigma(Y)) \mathbf{1}_{B_n(Y)\cap F_n(Y)}(g) \\
&= \sum_{g \in G_n} E\mu^{*n}(g) \left(\frac{d\lambda_g}{d\lambda}(Y) \mathbf{1}_{B_n(Y)\cap F_n(Y)}(g) \right) \\
&\leq \exp\{n(h_F - h) + 2n\varepsilon\}\, E \sum_{g \in G_n} \mathbf{1}_{B_n(Y)\cap F_n(Y)}(g) \\
&\leq \exp\{n(h_F - h) + 2n\varepsilon\}\, E|B_n(Y)| \\
&\leq \exp\{n(h_F - h) + 3n\varepsilon\}.
\end{aligned}
$$

Since this holds for any $\varepsilon > 0$, we must have $h_F = h$, because $h_F < h$ would imply that for all sufficiently small $\varepsilon > 0$,

$$\lim_{n\to\infty} P(B_n \cap F_n(Y) \cap G_n) = 0,$$

in contradiction to (12.2.14). Thus, Theorem 12.2.4 implies that the measure space $(\Upsilon, \mathcal{G}, \lambda)$ is a Poisson boundary. □

12.3 Poisson Boundaries of Lamplighter Walks†

Recall (cf. Section 3.4) that a *lamplighter random walk* is a random walk $(S_n, L_n)_{n\geq 0}$ on a restricted wreath product group $\mathbb{L} = \mathbb{Z}_2 \wr \Gamma$, where Γ is an infinite, finitely generated group. Elements of $\mathbb{L}$ are ordered pairs (x, ψ), where $x \in \Gamma$ and $\psi \in \oplus_\Gamma \mathbb{Z}_2$ is an element of the additive group of finitely supported *lamp configurations* on Γ; the identity element is $(1, \mathbf{0})$, where $\mathbf{0}$ is the configuration with all lamps off. The projection $(S_n)_{n\geq 0}$ of the lamplighter walk to the group Γ is called the *base random walk*. If the step distribution has finite support and the base random walk is transient, then by Proposition 3.4.2 the lamp configuration L_n stabilizes as $n \to \infty$, that is, for each site $x \in \Gamma$

$$L_\infty(x) := \lim_{n\to\infty} L_n(x) \tag{12.3.1}$$

exists almost surely.

Denote by $\mathbb{Z}_2^\Gamma$ the space of all *unrestricted lamp configurations* on Γ, that is, arbitrary functions $\psi : \Gamma \to \mathbb{Z}_2$. The space $\mathbb{Z}_2^\Gamma$ is compact and metrizable in the product topology, and the space $\oplus_\Gamma \mathbb{Z}_2$ of *restricted* configurations is dense in $\mathbb{Z}_2^\Gamma$. There is a natural action of the wreath product group $\mathbb{L}$ on $\mathbb{Z}_2^\Gamma$, defined by

$$(x, \varphi) \cdot \psi = \varphi + \sigma^x \psi. \tag{12.3.2}$$

Here $\sigma^x : \mathbb{Z}_2^\Gamma \to C$ is the left translation map on lamp configurations defined by

$$(\sigma^x \psi)(y) = \psi(x^{-1} y). \tag{12.3.3}$$

Proposition 12.3.1 *Let* $(X_n)_{n\geq 0} = ((S_n, L_n))_{n\geq 0}$ *be an irreducible random walk on the wreath product group* $\mathbb{L} = \mathbb{Z}_2 \wr \Gamma$, *with finitely supported step distribution* μ. *If the base random walk* $(S_n)_{n\geq 0}$ *is transient, then*

(i) the distribution λ *of* L_∞ *under* $P^{(1,\mathbf{0})}$ *is* μ*-stationary, and*
(ii) the pair $(\mathbb{Z}_2^\Gamma, \lambda)$ *is a* μ*-boundary.*

Proof Sketch. Let $\mathbf{X} = (X_n)_{n\geq 0} = (S_n, L_n)_{n\geq 0}$ and $\mathbf{X}' = (X_n')_{n\geq 0} = (S_n', L_n')_{n\geq 0}$ be independent copies of the lamplighter random walk, with increments ξ_n and

ξ'_n and limit configurations L_∞ and L'_∞, respectively. By definition, both L_∞ and L'_∞ have distribution λ. Consider the effect of prepending the increment ξ_1 to the increment sequence $(\xi'_n)_{n\geq 1}$: the resulting random walk path

$$X''_n := \xi_1 \xi'_1 \xi'_2 \cdots \xi_{n-1}$$

is nothing more than the random walk path $(X'_n)_{n\geq 0}$ translated to the new initial point ξ_1. Consequently, the limit configuration L''_∞ for this new random walk is the configuration $\xi_1 \cdot L'_\infty$. Since L'_∞ and L''_∞ both have distribution λ, the relation $L''_\infty = \xi_1 \cdot L'_\infty$ shows that λ is μ-stationary.

Next, fix any integer $m \geq 1$ and consider the random walk

$$X^m_n := \xi_1 \xi_2 \cdots \xi_m \xi'_1 \xi'_2 \cdots \xi'_{n-m}$$

obtained by translating the path $(X'_n)_{n\geq 0}$ to the initial point X_m. The limit configuration L^m_∞ for this random walk satisfies $L^m_\infty = X_m \cdot L'_\infty$. In particular, L^m_∞ is the configuration obtained by superposing the translated configuration $\sigma^{S_m} L'_\infty$ and the finite configuration L_m (cf. equation (12.3.2)). Since the base random walk $(S_n)_{n\geq 0}$ is transient, as $m \to \infty$ the probability that S_n enters any particular finite neighborhood of the group identity 1 after time m converges to 0; hence, in any finite neighborhood the lamp configuration $L^m_\infty = X_m \cdot L'_\infty$ eventually stabilizes at L_∞, that is,

$$\lim_{m\to\infty} X_m \cdot L'_\infty = L_\infty \quad \text{almost surely.}$$

This implies that with probability one, the measures $X_m \cdot \lambda$ converge weakly to the point mass at L_∞. □

Denote by $\delta_1 \in \oplus_\Gamma \mathbb{Z}_2$ the configuration with the lamp at $1 \in \Gamma$ on and all others off.

Theorem 12.3.2 *Let $(X_n)_{n\geq 0} = ((S_n, L_n))_{n\geq 0}$ be an irreducible lamplighter random walk on $\mathbb{L} = \mathbb{Z}_2 \wr \Gamma$. Assume that*

(A) the step distribution μ has finite support;
(B) support$(\mu) \subset \Gamma \times \{\mathbf{0}\} \cup \Gamma \times \{\delta_1\}$;
(C) the base random walk $(S_n)_{n\geq 0}$ has positive speed ℓ; and
(D) the base random walk $(S_n)_{n\geq 0}$ has a positive density α of cut points.

Then $(\mathbb{Z}_2^\Gamma, \lambda)$ is a Furstenberg-Poisson boundary for the random walk: in particular, every bounded harmonic function is of the form $u(x) = E^x f(L_\infty)$, for some bounded, Borel measurable function $f : \mathbb{Z}_2^\Gamma \to \mathbb{R}$.

This is a special case of a theorem of Erschler [38]. Hypothesis (B) is unnecessary, but allows for a simpler proof, as it ensures that at each step the random walker can only alter the state of the lamp at his current location. Cut points of a random walk $(S_n)_{n\geq 0}$ were defined in Section 2.3 (cf. Definition 2.3.1); these are

pairs (n, S_n) such that

$$\{S_0, S_1, \cdots, S_n\} \cap \{S_{n+1}, S_{n+2}, \cdots\} = \emptyset.$$

According to Theorem 2.3.2, cut points occur in any transient random walk with a (non-random) limiting frequency $\alpha \geq 0$, and Proposition 2.3.7 implies that for any symmetric random walk whose spectral radius is less than 1 the constant α is positive. The Carne-Varopoulos inequality implies (cf. Corollary 4.1.2 and Proposition 4.1.3) that if in addition a symmetric random walk with spectral radius less than 1 has finitely supported step distribution then its speed is also positive. Therefore, by Kesten's theorem (Theorem 5.1.5), if the base group Γ is nonamenable and the base random walk $(S_n)_{n\geq 0}$ is symmetric then under hypothesis (A) the hypotheses (C)–(D) will hold automatically.

The main idea behind the proof of Theorem 12.3.2 — that a positive density of cut points provides a means of verifying Kaimanovich's localization criterion (Corollary 12.2.5) — is due to Erschler. Hypothesis (C), like (B), is unnecessary; see [38]. However, the hypothesis that the lamplighter random walk is irreducible cannot be removed. This hypothesis, together with the hypothesis that the step distribution has finite support, ensures (cf. Theorem 3.4.4) that for an irreducible lamplighter random walk the number of lamps turned on at time n grows linearly with n, that is, there is a positive constant β such that

$$\lim_{n\to\infty} \frac{|\text{support}(L_n)|}{n} = \beta \quad \text{almost surely.} \tag{12.3.4}$$

Since the lamps that are eventually turned on must be at sites visited by the base random walk, the final lamp configuration L_∞ will (as we will show) provide enough information to allow a rough reconstruction of the lamplighter walk.

Proof of Theorem 12.3.2. The proof will be based on the localization criterion of Corollary 12.2.5. This criterion involves only the behavior of the random walk under the measure $P = P^{(1,\mathbf{0})}$, so we shall assume throughout that the initial point of the lamplighter random walk is the group identity $(1, \mathbf{0})$. In addition, because the step distribution of the lamplighter random walk is finitely supported, we may assume without loss of generality that the generating set of the group Γ (denote this by $\mathbb{A}^\Gamma$) has been chosen in such a way that the base random walk is nearest-neighbor.

We will first show that $\text{support}(L_\infty) = \{x \in \Gamma : L_\infty(x) = 1\}$ contains a positive limiting fraction of the points visited by the base random walk $(S_n)_{n\geq 0}$. As noted above, under hypothesis (B) the random walker at any step can only modify the state of the lamp at his current location; consequently, at any time n,

$$\text{support}(L_n) \subset \mathcal{R}_n := \{S_0, S_1, \cdots, S_n\}, \tag{12.3.5}$$

and at any cut point time τ, the lamp configuration L_τ is the restriction of the final lamp configuration L_∞ to the set of sites visited by time τ, that is,

$$\text{support}(L_\tau) = \text{support}(L_\infty) \cap \mathcal{R}_\tau. \tag{12.3.6}$$

By hypothesis (D), cut points occur with positive limiting frequency α. Denote the successive cut point times by $0 < \tau_1 < \tau_2 < \cdots$; then by Theorem 2.3.2,

$$\lim_{m\to\infty} \tau_m/m = 1/\alpha \quad \text{almost surely.} \tag{12.3.7}$$

Consequently, $\lim_{m\to\infty} \tau_m = \infty$, and so by (12.3.6) and (12.3.4),

$$\lim_{m\to\infty} \frac{|\text{support}(L_\infty) \cap \mathcal{R}_{\tau_m}|}{\tau_m} = \beta > 0 \tag{12.3.8}$$

with probability 1. Since the sets $\mathcal{R}_n$ are nondecreasing in n, and since $\lim_{m\to\infty} \tau_{m+1}/\tau_m = 1$ almost surely, by (12.3.7), it follows from (12.3.8) that with probability 1,

$$\lim_{n\to\infty} \frac{|\text{support}(L_\infty) \cap \mathcal{R}_n|}{n} = \beta > 0. \tag{12.3.9}$$

The relation (12.3.9) shows that the support of the final lamp configuration L_∞ contains a positive fraction of the sites visited by the base random walk; however, in general it will be impossible to deduce from L_∞ the order in which these sites were visited. This is where hypothesis (C), that the base random walk has positive speed ℓ, comes in. This hypothesis implies that for all sufficiently large n the random walker's location in the base group Γ must (with probability 1, for any fixed $\varepsilon > 0$) satisfy

$$n\ell - n\varepsilon < |S_n| < n\ell + n\varepsilon. \tag{12.3.10}$$

Because this holds for any $\varepsilon > 0$, it follows that with probability 1, for all large n,

$$S_{n'} \in \mathbb{B}^\Gamma_{n(\ell+2\varepsilon)} \quad \text{for all } n' \le n + n\varepsilon; \tag{12.3.11}$$

$$S_{n'} \notin \mathbb{B}^\Gamma_{n(\ell-2\varepsilon)} \quad \text{for all } n' \ge n - n\varepsilon; \quad \text{and} \tag{12.3.12}$$

$$S_{n'} \notin A_{n;2\varepsilon} \quad \text{for all } n' \notin [n - 4n\varepsilon, n + 4n\varepsilon]. \tag{12.3.13}$$

Here $\mathbb{B}^\Gamma_R$ denotes the ball of radius R in Γ centered at the group identity, and

$$A_{n;2\varepsilon} := \mathbb{B}^\Gamma_{n(\ell+2\varepsilon)} \setminus \mathbb{B}^\Gamma_{n(\ell-2\varepsilon)}.$$

Relations (12.3.11)–(12.3.12) imply that $\mathcal{R}_{n+n\varepsilon} \setminus \mathcal{R}_{n-n\varepsilon} \subset A_{n;2\varepsilon}$, so it follows from (12.3.9) that (for any $\varepsilon > 0$, with probability 1) for all sufficiently large n,

$$\text{support}(L_\infty) \cap A_{n;2\varepsilon} \neq \emptyset. \tag{12.3.14}$$

The restrictions (12.3.11)–(12.3.13) on the location of the base random walk make it possible to approximately locate both the base random walker and the lamp configuration at large cut point times. Consider first the lamp configuration. Since at each step the random walker can only modify the lamp at his current location, the relation (12.3.11) implies that with probability one, for all large n, the lamp configuration in the ball $\mathbb{B}^{\Gamma}_{n(\ell-2\varepsilon)}$ remains frozen after time $n-n\varepsilon$, and (12.3.12) implies that no lamps outside the ball $\mathbb{B}^{\Gamma}_{n(\ell+2\varepsilon)}$ are lit before time $n+n\varepsilon$. Hence, at any cut point time τ_m in the interval $[n-n\varepsilon, n+n\varepsilon]$, the lamp configuration must satisfy

$$\text{support}(L_{\tau_m}) \cap \mathbb{B}^{\Gamma}_{n(\ell-2\varepsilon)} = \text{support}(L_\infty) \cap \mathbb{B}^{\Gamma}_{n(\ell-2\varepsilon)} \quad \text{and} \tag{12.3.15}$$

$$\text{support}(L_{\tau_m}) \cap (\mathbb{B}^{\Gamma}_{n(\ell-2\varepsilon)})^c \subset \text{support}(L_\infty) \cap A_{n;2\varepsilon}, \tag{12.3.16}$$

for all sufficiently large n, with probability 1. Furthermore, by (12.3.13),

$$\text{support}(L_\infty) \cap A_{n;2\varepsilon} \subset \{S_{n-4n\varepsilon}, S_{n-4n\varepsilon+1}, \cdots, S_{n+4n\varepsilon}\}, \tag{12.3.17}$$

so with probability approaching 1 as $n \to \infty$, the set of possible lamp configurations that satisfy (12.3.15) and (12.3.16) has cardinality at most $2^{8n\varepsilon}$. Denote this set of configurations by $\mathcal{C}_{n;\varepsilon}(L_\infty)$.

Next, consider the location of the base random walk at a cut point time τ_m in the interval $[n-n\varepsilon, n+n\varepsilon]$. By (12.3.14), there is at least one site $x \in \text{support}(L_\infty)$ in the annulus $A_{n;2\varepsilon}$. Since the base random walk is nearest-neighbor, the random walker must at any time $n' \in [n-n\varepsilon, n+n\varepsilon]$ be somewhere in the ball $\mathbb{B}^{\Gamma}_{2n\varepsilon}(x)$. Thus, with probability 1, for all sufficiently large n, for any cut point time τ_m in the interval $[n-n\varepsilon, n+n\varepsilon]$,

$$S_{\tau_m} \in \mathcal{D}_{n;\varepsilon}(L_\infty) := \bigcup_{x \in \text{support}(L_\infty) \cap A_{n;2\varepsilon}} \mathbb{B}^{\Gamma}_{2n\varepsilon}(x). \tag{12.3.18}$$

Each ball $\mathbb{B}^{\Gamma}_{2n\varepsilon}(x)$ has cardinality $\leq A^{2n\varepsilon}$, where A is the cardinality of the generating set $\mathbb{A}^{\Gamma}$, and with probability approaching 1 as $n \to \infty$ the number of points x in the union (12.3.18) is at most $8n\varepsilon$, by (12.3.17); hence, with probability $\to 1$, the set $\mathcal{D}_{n;\varepsilon}(L_\infty)$ has cardinality $\leq 8n\varepsilon A^{2n\varepsilon}$.

Finally, define $\mathcal{Z}_{n;\varepsilon}(L_\infty)$ to be the set $\mathcal{D}_{n;\varepsilon}(L_\infty) \times \mathcal{C}_{n;\varepsilon}(L_\infty)$, provided this set has cardinality no larger than $8n\varepsilon A^{2n}\varepsilon \times 2^{8n\varepsilon}$, and $\mathcal{Z}_{n;\varepsilon}(L_\infty) = \emptyset$ otherwise. We have shown that for any $\varepsilon > 0$, with probability one, if n is sufficiently large and if n is a cut point time then $L_n \in \mathcal{C}_{n;\varepsilon}(L_\infty)$ and $S_n \in \mathcal{D}_{n;\varepsilon}(L_\infty)$. By Theorem 2.3.2 (specifically, relation (2.3.2)), the probability that n is a cut point time converges to $\alpha > 0$ as $n \to \infty$. Therefore,

$$\lim_{n\to\infty} P\left\{(S_n, L_n) \in \mathcal{Z}_{n;\varepsilon}(L_\infty)\right\} = \alpha > 0.$$

Since $\varepsilon > 0$ is arbitrary, this proves that Kaimanovich's localization criterion is satisfied, and so it follows by Corollary 12.2.5 that the pair $(\mathbb{Z}_2^\Gamma, \lambda)$ is a Poisson boundary. □

12.4 Existence of Poisson Boundaries

The construction of a Furstenberg-Poisson boundary requires a suitable compact metric space on which the group Γ acts. Natural candidates for such metric space are spaces of functions on Γ with the product topologies (i.e., the topology of pointwise convergence); the group Γ acts naturally on any such function space by left translations. We will use the space $\mathcal{K}$ of positive harmonic functions that take the value 1 at the group identity:

$$\mathcal{K} := \{\text{harmonic functions } h \,:\, h(1) = 1 \;\text{ and }\; h > 0\} \tag{12.4.1}$$

Proposition 9.2.7 implies that for ν-almost every $\mathbf{x} \in \Gamma^\infty$ the likelihood ratio function

$$g \mapsto L_g(\mathbf{x}) := \left(\frac{d\nu_g}{d\nu}\right)(\mathbf{x})$$

is an element of $\mathcal{K}$, and Theorem 9.1.5 implies that every bounded harmonic function u has an essentially unique representation $u(g) = E_\nu(f L_g)$. Thus, the choice of the space $\mathcal{K}$ is a natural starting point for the construction of a Poisson boundary.

By the Harnack inequalities for positive harmonic functions (cf. Exercise 6.2.5 in Chapter 6), every function $h \in \mathcal{K}$ satisfies the inequalities

$$C_g^- \le h(g) \le C_g^+ \tag{12.4.2}$$

where $C_g^- = \max_{n\ge 0} \mu^{*n}(g^{-1})$ and $C_g^+ = 1/\max_{n\ge 0} \mu^{*n}(g)$; since the random walk is, by hypothesis, irreducible, the constants $C_g^\pm$ are positive and finite. Hence, the space $\mathcal{K}$ can be viewed as a closed — and therefore compact — subset of $\prod_{g\in\Gamma}[C_g^-, C_g^+]$ in the product topology. Because the index set Γ is countable, the product topology is metrizable: for instance, if $(g_n)_{n\ge 1}$ is an enumeration of Γ, the metric

$$d(h_1, h_2) = \sum_{n=1}^{\infty} |h_1(g_n) - h_2(g_n)|/(2^n C_{g_n}^+)$$

induces the product topology. The group Γ acts on $\mathcal{K}$ by left translation as follows: for any $g \in \Gamma$ and $h \in \mathcal{K}$,

$$(g \cdot h)(x) = \frac{h(g^{-1}x)}{h(g^{-1})}. \tag{12.4.3}$$

Exercise 12.4.1 Check that this is a Γ-action, that is, $g_1 \cdot (g_2 \cdot h) = (g_1 g_2) \cdot h$ for every $h \in \mathcal{K}$ and all $g_1, g_2 \in \Gamma$.

For any $g \in \Gamma$ the mapping $h \mapsto g \cdot h$ is invertible and continuous relative to the product topology on $\mathcal{K}$, so (12.4.3) does in fact define a group homomorphism from Γ to the group of homeomorphisms of $\mathcal{K}$.

Lemma 12.4.2 *For each* $\mathbf{x} \in \Gamma^\infty$ *define* $\Lambda(\mathbf{x}) \in \mathcal{K}$ *to be the harmonic function whose value at any* $g \in \Gamma$ *is*

$$(\Lambda(\mathbf{x}))(g) = L_g(\mathbf{x}) = \frac{d\nu_g}{d\nu}(\mathbf{x}), \tag{12.4.4}$$

provided this is harmonic, and define $\Lambda(\mathbf{x}) \equiv 1$ *otherwise. The resulting function* $\Lambda : \Gamma^\infty \to \mathcal{K}$ *is an* equivariant boundary map *for the* Γ*-action (12.4.3). In particular,* Λ *is* $\mathcal{I}$*-measurable (relative to the Borel* σ*-algebra on* $\mathcal{K}$*) and satisfies the intertwining relation*

$$g \cdot \Lambda(\mathbf{x}) = \Lambda(g \cdot \mathbf{x}) \quad \textit{for all} \;\; g \in \Gamma, \;\; \mathbf{x} \in \Gamma^\infty. \tag{12.4.5}$$

Proof. Measurability follows directly from the $\mathcal{I}$-measurability of the likelihood ratios $L_g(\mathbf{x})$. Relation (12.4.5) follows by the transformation rule for Radon-Nikodym derivatives (see Exercise A.8.7 of the Appendix). This implies that for any $g_1, g_2 \in \Gamma$ and $\mathbf{x} \in \Gamma^\infty$,

$$\frac{d\nu_{g_1 g_2}}{d\nu_{g_1}}(g_1 \cdot \mathbf{x}) = \frac{d(\nu_{g_2} \circ g_1^{-1})}{d(\nu \circ g_1^{-1})}(g_1 \cdot \mathbf{x}) = \frac{d\nu_{g_2}}{d\nu}(\mathbf{x});$$

consequently, by (12.4.3),

$$g_3 \cdot \Lambda(\mathbf{x})(g_4) = \frac{\Lambda(\mathbf{x})(g_3^{-1} g_4)}{\Lambda(\mathbf{x})(g_3^{-1})} = \frac{d\nu_{g_3^{-1} g_4}}{d\nu_{g_3^{-1}}}(\mathbf{x}) = \frac{d\nu_{g_4}}{d\nu}(g_3 \cdot \mathbf{x}) = \Lambda(g_3 \cdot \mathbf{x})(g_4).$$

□

The intertwining relation (12.4.5) extends to the induced Γ-actions on measures (cf. equation (12.1.8)): in particular, (12.4.5) implies that for any $g_1, g_2 \in \Gamma$,

$$g_1 \cdot (\nu_{g_2} \circ \Lambda^{-1}) = (g_1 \cdot \nu_{g_2}) \circ \Lambda^{-1} = \nu_{g_1 g_2} \circ \Lambda^{-1}. \tag{12.4.6}$$

Furthermore, by Proposition 12.1.8, the induced measure $\lambda := \nu \circ \Lambda^{-1}$ on $(\mathcal{K}, \mathcal{B}_{\mathcal{K}})$ is μ-stationary, so the pair $(\mathcal{K}, \lambda)$ is a μ-space, in the sense of Definition 11.1.1. To

show that it is a μ*-boundary*, we must verify that for any $g \in \Gamma$, with P^g-probability one the random measures $X_n \cdot \lambda$ converge weakly to point masses δ_Y.

Lemma 12.4.3 *For every $g \in \Gamma$, with P^g-probability one the measures $X_n \cdot \lambda$ converge weakly as $n \to \infty$ to the point mass at $\Lambda(\mathbf{X})$, that is, for every continuous function $f : \mathcal{K} \to \mathbb{R}$,*

$$\lim_{n\to\infty} \int f(X_n \cdot h)\, d\lambda(h) = f(\Lambda(\mathbf{X})) \quad P^g - \textit{almost surely}. \tag{12.4.7}$$

Proof. Without loss of generality, assume that $P^x = \hat{P}^x$ is the law of the random walk with initial point x and that $X_n = \hat{X}_n : \Gamma^\infty \to \Gamma$ are the coordinate projections. The pullback $f \circ \Lambda$ is a bounded, $\mathcal{I}$-measurable function on Γ^∞, so by Proposition 9.1.4 the function $u(g') := E_\nu(f \circ \Lambda) L_{g'} = E_{\nu_{g'}}(f \circ \Lambda)$ is bounded and harmonic. Therefore, by Theorem 9.1.5,

$$\lim_{n\to\infty} u(X_n) = (f \circ \Lambda)(\mathbf{X}) \quad P^g - \text{almost surely}.$$

By the intertwining relation (12.4.5) and the definition $\lambda = \nu \circ \Lambda^{-1}$,

$$\begin{aligned}
\int f(X_n \cdot h)\, d\lambda(h) &= \int f(X_n \cdot \Lambda(\mathbf{x}))\, d\nu(\mathbf{x}) \\
&= \int f(\Lambda(X_n \cdot \mathbf{x}))\, d\nu(\mathbf{x}) \\
&= \int (f \circ \Lambda)(\mathbf{x})\, d\nu_{X_n}(\mathbf{x}) \\
&= u(X_n) \longrightarrow (f \circ \Lambda)(\mathbf{X}) \quad P^g - \text{almost surely}.
\end{aligned}$$

□

Proof of Theorem 12.1.4. Let $\mathcal{K}^*$ be the closure of the set $\Lambda(\Gamma) = \{\Lambda(g) : g \in \Gamma\}$ in $\mathcal{K}$. The intertwining relation (12.4.5) implies that the action (12.4.3) of Γ on $\mathcal{K}$ leaves the set $\mathcal{K}^*$ invariant, and also implies that the resulting action of Γ on $\mathcal{K}^*$ is *transitive*. Lemma 12.4.3 implies that the pair $(\mathcal{K}^*, \lambda)$ with the Γ-action (12.4.3) is a μ-boundary, with associated equivariant boundary map $Y = \Lambda$. It remains to prove that $(\mathcal{K}^*, \lambda)$ is a Poisson boundary.

By Corollary 12.1.10, the Banach spaces $L^\infty(\mathcal{K}^*, \mathcal{B}_{\mathcal{K}^*}, \lambda)$ and $L^\infty(\Gamma^\infty, \sigma(Y), \nu)$ are linearly isomorphic, and by equations (9.1.6) and (12.1.1), elements of these Banach spaces correspond to bounded harmonic functions. Consequently, to complete the proof it suffices to show that $(\Gamma^\infty, \sigma(Y), \nu)$ is a Poisson boundary. For this it is enough, by Corollary 12.1.12, to show that the likelihood ratios L_g are all measurable with respect to $\sigma(Y)$. But the likelihood ratio L_g is just the evaluation of Λ at the point g, that is, $L_g = e_g \circ \Lambda$ where $e_g : \mathcal{K} \to \mathbb{R}$ is the function

$e_g(h) = h(g)$; since $e_g : \mathcal{K} \to \mathbb{R}$ is continuous, it is also Borel measurable, and so it follows that L_g is measurable with respect to $\Lambda^{-1}(\mathcal{B}_\mathcal{Y}) = \sigma(Y)$. □

The existence of Furstenberg-Poisson boundaries, together with Theorem 10.1.1, provides a painless proof of the converse of Proposition 11.2.7.

Proposition 12.4.4 *If every compact* Γ*-space admits a* Γ*-invariant Borel probability measure then* Γ *is amenable.*

Proof. Let μ be a symmetric, finitely-supported probability measure on Γ whose support generates Γ, and let $(\mathcal{Y}, \lambda)$ be a Furstenberg-Poisson boundary for the random walk $(X_n)_{n\geq 0}$ with step distribution μ. By hypothesis, the space $\mathcal{Y}$ carries a Γ-invariant Borel probability measure λ'. We will prove that under this hypothesis, the space $\mathcal{Y}$ consists of just a single point; this will imply that there are no non-constant bounded harmonic functions. It will then follow from Theorem 10.1.1 and Corollary 5.1.7 that Γ is amenable.

Suppose first that the space $\mathcal{Y}$ is infinite. Then by Proposition 11.3.4, the measure λ is nonatomic and charges every nonempty open subset of $\mathcal{Y}$, and by Proposition 11.4.13, λ is the *unique* μ-stationary measure. Since any Γ-invariant probability measure is μ-stationary, it follows that $\lambda' = \lambda$. Consequently, for any group element $g \in \Gamma$, we have $\lambda = g \cdot \lambda$. But because the space $(\mathcal{Y}, \lambda)$ is a μ-boundary, with probability one the measures $X_n \cdot \lambda$ converge to a point mass. This implies that λ must be a point mass, contradicting the fact that it is nonatomic.

Therefore, the space $\mathcal{Y}$ is *finite.* Since the action of Γ on $\mathcal{Y}$ is, by definition of a Furstenberg-Poisson boundary, *transitive*, Exercise 11.3.5 implies that the uniform distribution on $\mathcal{Y}$ is the unique μ-stationary distribution, so $\lambda' = \lambda =$uniform. But once again with probability one the measures $X_n \cdot \lambda$ converge to a point mass, and so λ must be a point mass. The only possibility is that the space $\mathcal{Y}$ is a singleton. □

Additional Notes. The concept of a Poisson boundary originates in the work of Furstenberg, in particular, the article [47]. His original definition requires that the underlying space be a μ-boundary, whereas, in modern terminology, a Poisson boundary is a purely measure-theoretic object. To differentiate, we have used the term *Furstenberg-Poisson boundary* for the space studied in [47]. The existence of Furstenberg-Poisson boundaries was a central result of [47]; the proof relied on the fact that the space of bounded, μ-harmonic functions is a *Banach algebra.* The proof laid out in Section 12.4, which avoids the use of Banach algebra theory, is not only more elementary but also seems more natural.

The characterization of Poisson boundaries by the equality of Furstenberg and Avez entropies is one of the main results of the important article [68] (cf. their Theorem 3.2), but is prefigured by a similar result for the groups $SL(n, \mathbb{Z})$ in [47].

Kaimanovich's localization criterion (Corollary 12.2.5) has become one of the most useful tools for identifying Poisson boundaries. The examples presented in Section 12.3 and Section 13.5 are two of the more prominent examples of its use. Theorem 12.3.2 is a special case of a theorem of A. Erschler [38], who also proved the corresponding theorem for random walks on the lamplighter groups $\mathbb{Z}_2 \wr \mathbb{Z}^d$ for $d \geq 5$. Symmetric nearest-neighbor random walks on the groups $\mathbb{Z}_2 \wr \mathbb{Z}^d$ with $d = 1$

or 2 have speed 0 (cf. Proposition 1.7.4 for an instance of this when $d = 1$), and consequently, by Theorem 10.1.1, have trivial Poisson boundaries. Lyons & Peres [92] have proved that Erschler's result for the lamplighter groups $\mathbb{Z}_2 \wr \mathbb{Z}^d$ extends to the intermediate dimensions $d = 3, 4$.

Chapter 13
Hyperbolic Groups

A recurring theme of these lectures has been the interplay between the geometry of a group's Cayley graph(s) and the behavior of random walks on the group. In this chapter, we shall study a large class of groups, the *hyperbolic groups*, first introduced by Gromov [59], whose geometry forces certain regularities on random walk paths.

Definition 13.0.1 A finitely generated group Γ is *hyperbolic* in the sense of Gromov if for some (any) finite, symmetric generating set $\mathbb{A}$ the Cayley graph $G_{\Gamma;\mathbb{A}}$ is a *hyperbolic metric space*, that is, a geodesic metric space such that for some $\delta \geq 0$, every geodesic triangle is δ*-thin.*

To make sense of this definition we must explain what a *geodesic metric space* is and what it means for a triangle to be δ-thin, and we must show that hyperbolicity is independent of the choice of generating set. This we will do in Section 13.1.

13.1 Hyperbolic Metric Spaces

A *rectifiable path* in a metric space $(\mathcal{X}, d)$ is a continuous path $\beta : [0, L] \to \mathcal{X}$ (or its image $\beta[0, L]$) such that

$$\text{arc length}(\beta) := \sup \sum_{i=0}^{n-1} d(\beta(t_i), \beta(t_{i+1})) < \infty, \tag{13.1.1}$$

where the supremum is over all finite partitions of the interval $[0, L]$. Any rectifiable path β can be parametrized by arc length, and we shall assume without further comment that β is given this parametrization; thus, for any $0 \leq L' \leq L$, the arc length of the restriction $\beta \restriction [0, L']$ is L'. A *geodesic* is an *isometry* $\gamma : J \to \mathcal{X}$ of an interval J of the real line $\mathbb{R}$, that is,

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_13

$$d(\gamma(s), \gamma(t)) = |t - s| \quad \text{for all} \quad s, t \in J. \tag{13.1.2}$$

Such a path is a geodesic *segment, ray,* or *line* according as J is a finite interval, halfline, or $\mathbb{R}$, respectively. In the first two cases, we shall assume that the parametrization is such that the interval J has 0 as its left endpoint.

A metric space $(\mathcal{X}, d)$ is a *geodesic space* if for any two points $x, y \in \mathcal{X}$ there is a geodesic segment with endpoints x and y. (This geodesic segment need not be unique: for example, any two antipodal points on the unit circle are connected by two geodesic segments.) Any Cayley graph — or, more generally, any connected graph G with vertex set V and edge set $\mathcal{E}$ — can be viewed as a geodesic space by adopting the convention that each edge $e \in \mathcal{E}$ is a Euclidean line segment of length 1. The usual graph distance on the vertex set (the word metric, if the graph is a Cayley graph) then extends in the obvious way to a metric on the space $V \cup \mathcal{E}$, and by virtue of the connectivity of G the resulting metric space is geodesic.

A *geodesic triangle* in a geodesic space $(\mathcal{X}, d)$ is a set of three geodesic segments connecting three points of $\mathcal{X}$ in pairs. Similarly, a *geodesic quadrilateral* is a set of four geodesic segments connecting four points in sequence.

Definition 13.1.1 (Hyperbolic Metric Space) A geodesic space $(\mathcal{X}, d)$ is δ-hyperbolic, for some real $\delta \geq 0$, if every geodesic triangle is δ-*thin*, that is, each side is contained in the δ-neighborhood of the union of the other two sides. The space $(\mathcal{X}, d)$ is *hyperbolic* if it is δ-hyperbolic for some $\delta \geq 0$.

Obviously, if a geodesic space $(\mathcal{X}, d)$ is δ-hyperbolic then it is δ'-hyperbolic for any $\delta' \geq \delta$. Thus, when necessary, we can always take the parameter δ to be an integer.

Example 13.1.2 The simplest examples of hyperbolic metric spaces are the infinite homogeneous trees $\mathbb{T}_d$. Any tree $\mathbb{T}_d$ is 0-hyperbolic, because each side of any geodesic triangle is contained in the union of the other two sides.

Example 13.1.3 † The *hyperbolic plane* $\mathbb{H}$ is the model hyperbolic space from which Definition 13.1.1 was abstracted. The space $\mathbb{H}$ has several equivalent descriptions, of which the *Poincaré disk model* is most useful for our purposes. In this model, $\mathbb{H}$ is identified with the open unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ of the complex plane, and given the metric

$$d(z_1, z_2) = \inf \int_\gamma \frac{2|dz|}{1 - |z|^2}, \tag{13.1.3}$$

where the inf is over all paths γ from z_1 to z_2. See, for instance, [8] for a thorough discussion. Geodesic lines in this metric are arcs of Euclidean circles with endpoints on the unit circle $\partial\mathbb{D} = \{z \in \mathbb{C} : |z| = 1\}$ that meet the unit circle at (Euclidean) right angles. The isometries of $\mathbb{H}$ are the linear fractional transformations of $\mathbb{C}$ that leave the unit circle (and hence also the disk $\mathbb{D}$) invariant.

Following is a sketch of a proof that the space $\mathbb{H}$ is hyperbolic in the sense of Definition 13.1.1. First, it suffices to show that for some $\delta < \infty$, the interior of every

geodesic triangle with endpoints on the "ideal boundary" $\partial\mathbb{D}$ (that is, the union of three geodesic lines that meet, in pairs, at three points $\xi_1, \xi_2, \xi_3 \in \partial\mathbb{D}$) is δ-thin. This is because any geodesic triangle in $\mathbb{D}$ can be enclosed in a geodesic triangle with "vertices" on $\partial\mathbb{D}$. For any triple $\xi_1, \xi_2, \xi_3 \in \partial\mathbb{D}$, there is an isometry of $\mathbb{H}$ that maps ξ_1, ξ_2, ξ_3 to $\zeta^0, \zeta^1, \zeta^2$, where $\zeta = e^{2\pi}i/3$; consequently, it suffices to prove that the geodesic triangle with vertices $\zeta^0, \zeta^1, \zeta^2$ is δ-thin, for some $\delta > 0$. For this it suffices to show that the hyperbolic distance between corresponding points on the two sides meeting at the boundary point $1 = \zeta^0$ decays to 0. This is most easily accomplished by mapping $\mathbb{H}$ isometrically to the upper half plane, where the corresponding Riemannian metric is $ds = |dz|/\mathrm{Im}(z)$, by a linear fractional transformation mapping 1 to ∞. Such a transformation takes the two geodesic sides meeting at $1 = \zeta^0$ to two vertical lines $\mathrm{Re}(z) = x_1$ and $\mathrm{Re}(z) = x_2$; the distances between corresponding points $x_1 + iy$ and $x_2 + iy$ on these lines is easily seen to decay to zero as $y \to \infty$.

Exercise 13.1.4 (Geodesic Quadrilaterals) Let Q be a geodesic quadrilateral in a δ-hyperbolic space such that Q has vertices A, B, C, D and sides $\alpha_1 = AB$, $\alpha_2 = BC$, $\alpha_3 = CD$, and $\alpha_4 = DA$. Show that if $d(\alpha_1, \alpha_3) > 2\delta$ then every point x on α_2 such that $\min(d(x, \alpha_1), d(x, \alpha_3)) > 2\delta$ is within distance 2δ of a point on α_4.

Exercise 13.1.5 (Geodesic Projections) Let γ be a geodesic in a δ-hyperbolic space $(\mathcal{X}, d)$, and let $x \in \mathcal{X}$ be a point not on γ.

(a) Show that there is at least one point y on γ such that $d(x, y) = \min_{z\in\gamma} d(x, z)$. Call any such point a *geodesic projection* of x onto γ.
NOTE: The existence of geodesic projections does not require hyperbolicity; it holds in any geodesic space.

(b) Let y be a geodesic projection of x onto γ. Show that for any other point z on γ,

$$d(x, z) \geq d(x, y) + d(y, z) - 4\delta, \tag{13.1.4}$$

and conclude that for any two geodesic projections $y, y' \in \gamma$ of x onto γ,

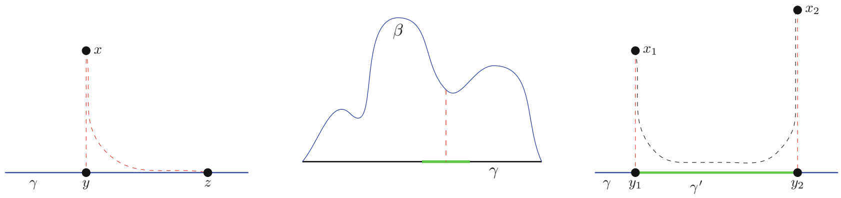
$$d(y, y') \leq 4\delta. \tag{13.1.5}$$

(c) Let β be a continuous path in $\mathcal{X}$ whose endpoints y, y' lie on the geodesic segment γ. Show that every sub-segment $\gamma' = y''y'''$ of γ with length $d(y'', y''') > 4\delta$ contains a geodesic projection of some point $x \in \beta$ onto γ.

(d) Let $x_1, x_2 \in \mathcal{X}$ be distinct points not on γ and let $y_1, y_2 \in \gamma$ be geodesic projections of x_1, x_2 onto γ, respectively. Show that each point on the geodesic segment $\gamma' = y_1y_2 \subset \gamma$ not within distance 4δ of either y_1 or y_2 must be within distance 2δ of any geodesic segment x_1x_2 with endpoints x_1, x_2.

HINTS: For part (b), show that any geodesic segment xz from x to z must stay within distance δ of any geodesic segment xy from x to y until it enters the δ-neighborhood of γ, and so must first enter this neighborhood at a point w within distance 2δ of y.

For part (d), use the hypothesis that y_i is a geodesic projection of x_i to show that any point y on y_1y_2 not within distance 4δ of y_1 or y_2 cannot be within distance 2δ of either x_1y_1 or x_2y_2, where x_iy_i is any geodesic segment connecting x_i and y_i.



(b) (c) (d)

Hyperbolicity of a geodesic metric space is, in a certain sense, a *coarse* property of the metric. To formulate this precisely, we will need the notion of a *quasi-isometry*.

Definition 13.1.6 (Quasi-Isometric Mapping) Fix real numbers $K \geq 1$ and $\varepsilon \geq 0$. A (not necessarily continuous) mapping $T : \mathcal{X} \to \mathcal{Y}$ between two metric spaces $(\mathcal{X}, d_X)$ and $(\mathcal{Y}, d_Y)$ is a (K, ε)-*quasi-isometric mapping* if for any two points $x_1, x_2 \in \mathcal{X}$,

$$K^{-1}d_X(x_1, x_2) - \varepsilon \leq d_Y(Tx_1, Tx_2) \leq K d_X(x_1, x_2) + \varepsilon. \tag{13.1.6}$$

A quasi-isometric mapping T is a *quasi-isometry* if it is *quasi-surjective*, that is, there exists $C < \infty$ such that every point $y \in \mathcal{Y}$ is within distance C of the image $T(\mathcal{X})$.

Exercise 13.1.7 Let Γ be a finitely generated group with finite, symmetric generating sets $\mathbb{A}$, $\mathbb{A}'$, and let $G = G_{\Gamma;\mathbb{A}}$ and $G' = G_{\Gamma;\mathbb{A}'}$ be the corresponding Cayley graphs. Show that the identity mapping $i : \Gamma \to \Gamma$ extends to a quasi-isometry $T : G \to G'$. (In other words, show that i can be extended to the points on the edges of G in such a way that the resulting mapping is a quasi-isometry.)

Proposition 13.1.8 *Let $(\mathcal{X}, d_X)$ and $(\mathcal{Y}, d_Y)$ be geodesic spaces. If $(\mathcal{Y}, d_Y)$ is a hyperbolic space and $\phi : \mathcal{X} \to \mathcal{Y}$ is a quasi-isometry then $\mathcal{X}$ is also a hyperbolic space.*

In view of Exercise 13.1.7, this implies that hyperbolicity of a finitely generated group does not depend on the choice of generating set. We defer the proof of Proposition 13.1.8 to Section 13.2.

The preservation of hyperbolicity under quasi-isometry leads to a useful tool, the *Milnor-Schwarz Lemma*, for identifying hyperbolic groups. Formulation of this

lemma requires several auxiliary notions. Recall that a metric space is *proper* if it has the *Heine-Borel* property, that is, every closed, bounded subset is compact. If a group Γ acts on a metric space $\mathcal{X}$ by homeomorphisms, the action is *discontinuous* if for every compact subset $C \subset \mathcal{X}$ the set $\{g \in \Gamma : g \cdot C \cap C \neq \emptyset\}$ is finite. The action is *cocompact* if the quotient space $\mathcal{X}/\Gamma$ (that is, the identification space for the equivalence relation $x_1 \sim x_2$ if and only if $x_2 \in \Gamma \cdot x_1$) is compact. If $(\mathcal{X}, d_X)$ is proper, then a group action of Γ by isometries of $\mathcal{X}$ is cocompact if and only if there exists a real number $r < \infty$ such that for any two points $x, y \in \mathcal{X}$ there exists $g \in \Gamma$ such that $d_X(g \cdot x, y) < r$.

Proposition 13.1.9 (Milnor-Schwarz) *If a group Γ acts discontinuously and cocompactly by isometries on a proper, geodesic metric space $(\mathcal{X}, d_X)$, then Γ is finitely generated, and for any $x \in \mathcal{X}$ the mapping $g \mapsto g \cdot x$ of Γ into $\mathcal{X}$ extends to a quasi-isometry of any Cayley graph of Γ to $\mathcal{X}$. Consequently, if $\mathcal{X}$ is a hyperbolic space then Γ is a hyperbolic group.*

The Cayley graph $G_{\Gamma;\mathbb{A}}$ of a finitely generated group Γ with finite, symmetric generating set $\mathbb{A}$ is a proper geodesic space, because for any $r > 0$ the ball of radius r centered at the group identity contains only finitely many vertices and edges. Moreover, the natural action of the group on its Cayley graph by left multiplication is discontinuous and cocompact. If H is a finite-index subgroup of Γ then the induced action of H on $G_{\Gamma;\mathbb{A}}$ is discontinuous and cocompact; consequently, if H is hyperbolic then so is Γ. Thus, the group $SL(2,\mathbb{Z})$ is hyperbolic, as it has a free group as a finite-index subgroup. Another noteworthy special case is that where the metric space $\mathcal{X} = \mathbb{H}$ is the hyperbolic plane. A discontinuous group of isometries of the hyperbolic plane is called a *Fuchsian group*; Proposition 13.1.9 implies that every cocompact Fuchsian group is hyperbolic.

Proof of Proposition 13.1.9. The proof that Γ is finitely generated mimics (and extends) the proof that a finite-index subgroup of a finitely generated group is itself finitely generated — see Exercise 1.2.11. Fix a reference point $x_* \in \mathcal{X}$. Since Γ acts cocompactly on $\mathcal{X}$, there exists $1 < r < \infty$ so large that every point of $\mathcal{X}$ lies within distance r of some point $g \cdot x_*$ in the orbit of x_*. Set

$$\mathbb{A} = \left\{ g^{\pm 1} \in \Gamma : g \cdot \mathbb{B}^{\circ}_{2r}(x_*) \cap \mathbb{B}^{\circ}_{2r}(x_*) \neq \emptyset \right\},$$

where $\mathbb{B}^{\circ}_{2r}(x)$ denotes the *open* ball of radius $2r$ centered at x. This set is finite, because Γ acts discontinuously, and it is symmetric by definition. We will argue that $\mathbb{A}$ is a generating set for Γ.

Fix $g \in \Gamma$. Since the metric space $\mathcal{X}$ is geodesic, there is a geodesic segment γ connecting x_* and $g \cdot x_*$. This geodesic segment passes successively through a finite sequence $\left\{ g_i \cdot \mathbb{B}^{\circ}_{2r}(x_*) \right\}_{1 \leq i \leq I}$ of translates of the ball $\mathbb{B}^{\circ}_{2r}(x_*)$ such that $g_0 = 1$ and $g_I = g$. Because these translates are open, each successive pair have nontrivial intersection, that is, $g_i \cdot \mathbb{B}^{\circ}_{2r}(x_*) \cap g_{i+1} \cdot \mathbb{B}^{\circ}_{2r}(x_*) \neq \emptyset$. But since the mappings $x \mapsto g \cdot x$ are bijective (being isometries) this implies that $\mathbb{B}^{\circ}_{2r}(x_*) \cap (g_i^{-1} g_{i+1}) \cdot$

$\mathbb{B}^{\circ}_{2r}(x_*) \neq \emptyset$, and so $g_i^{-1} g_{i+1} := a_i \in \mathbb{A}$. Therefore,

$$g = a_1 a_2 \cdots a_I.$$

This proves that $\mathbb{A}$ is a generating set for Γ.

It remains to prove that for any $x_* \in \mathcal{X}$ the mapping $T : \Gamma \to \mathcal{X}$ defined by $g \mapsto g \cdot x_*$ extends to a quasi-isometry of the Cayley graph $G_{\Gamma;\mathbb{A}}$ to $\mathcal{X}$. (It will then follow by Exercise 13.1.7 that *any* Cayley graph of Γ is quasi-isometric to $\mathcal{X}$.) The mapping T is quasi-surjective, because every point of $\mathcal{X}$ is within distance r of some point in the orbit $\Gamma \cdot x_*$, so what we must show is that T satisfies the inequalities (13.1.6) for some $K \geq 1$ and $\varepsilon \geq 0$.

Denote by d_w the natural extension to the Cayley graph of the word metric on Γ with respect to the generating set $\mathbb{A}$. Fix $g \in \Gamma$, and as above let γ be a geodesic segment with endpoints x_* and $g \cdot x_*$. Let $x_0 = x_*, x_1, \cdots, x_m$ be points on γ such that $d_X(x_i, x_{i+1}) = 1$ and $0 \leq d_X(x_m, g \cdot x_*) < 1$; thus, $m \leq d_X(x_*, g \cdot x_*) < m+1$. Each point x_i lies within distance r of a point $g_i \cdot x_*$ in the orbit of x_*, and so for each i the group element $g_i^{-1} g_{i+1}$ is an element of $\mathbb{A}$, as is $g_m^{-1} g$. Therefore,

$$d_w(1, g) \leq m + 1 \leq d_X(x_*, g \cdot x_*) + 1. \tag{13.1.7}$$

Conversely, let $g = a_1 a_2 \cdots a_m$ where each $a_i \in \mathbb{A}$. Let β be the path in $\mathcal{X}$ from x_* to $g \cdot x_*$ gotten by concatenating geodesic segments from $g_i \cdot x_*$ to $g_{i+1} x_*$, where $g_i = a_1 a_2 \cdots a_i$. Each of the geodesic segments has length $\leq 4r$, so the arc length of β is bounded by $4rm$. Hence,

$$d_X(x_*, g \cdot x_*) \leq 4r d_w(1, g). \tag{13.1.8}$$

The inequalities (13.1.7)–(13.1.8) and the fact that elements of Γ act as isometries on $\mathcal{X}$ imply that the restriction of T to Γ is a (K, ε)-quasi-isometric mapping, with $K = 4r + 1$ and $\varepsilon = 1$. The extension to the full Cayley graph (that is, to points on the edges) follows routinely (with possibly larger values of K and ε). □

13.2 Quasi-Geodesics

The proof of Proposition 13.1.8 will turn on the fact that quasi-isometric images of geodesic segments in a hyperbolic metric space $\mathcal{Y}$ are closely approximated by geodesic segments. Closeness is measured by *Hausdorff distance*: for any two subsets A, B of a metric space $(\mathcal{X}, d)$, the Hausdorff distance between A and B is defined to be the maximum of $\sup_{x \in A} d(x, B)$ and $\sup_{y \in B} d(y, A)$.

Exercise 13.2.1 Check that Hausdorff distance is a *metric* on the set of compact subsets of a metric space.

Definition 13.2.2 A (K, ε)-*quasi-geodesic* in a metric space $(\mathcal{X}, d)$ is a (K, ε)-quasi-isometric mapping $\beta : J \to \mathcal{X}$ of a closed interval $J \subset \mathbb{R}$, or alternatively the

image of such a mapping. A path is quasi-geodesic if it is (K, ε)-quasi-geodesic for some $K \geq 1$ and $\varepsilon \geq 0$.

A composition of quasi-isometries is a quasi-isometry, and in particular the composition of a (K, ε)-quasi-isometry with an isometry is again a (K, ε)-quasi-isometry. Consequently, the image of any geodesic in a metric space $\mathcal{X}$ under a (K, ε)-quasi-isometry $f : \mathcal{X} \to \mathcal{Y}$ is a (K, ε)-quasi-geodesic.

Quasi-geodesy is a *coarse* geometric property of a path; it does not depend on behavior over short distances (of order ε or smaller). In particular, a quasi-geodesic need not be a continuous path. However, if the ambient metric space $(\mathcal{X}, d)$ is *geodesic*, then any quasi-geodesic can be closely approximated by a continuous, *piecewise geodesic* quasi-geodesic: more precisely, for any $K \geq 1$ and $\varepsilon > 0$ there exist constants $C(K, \varepsilon) < \infty$ and K', ε' such that any (K, ε)-quasi-geodesic segment is within Hausdorff distance C of a continuous, piecewise geodesic, (K', ε')-quasi-geodesic segment.

Exercise 13.2.3 Prove this.

A key to the coarse geometry of hyperbolic spaces is the fact that quasi-geodesics in any hyperbolic space, regardless of their lengths, are closely shadowed by geodesics:

Proposition 13.2.4 (Morse) *If $(\mathcal{X}, d)$ is a δ-hyperbolic space, then for any $K \geq 1$ and $\varepsilon \geq 0$ there exists $C = C(K, \varepsilon, \delta) < \infty$ such that for any (K, ε)-quasi-geodesic segment β in $\mathcal{X}$ and any geodesic segment γ with the same endpoints, the Hausdorff distance between β and γ is less than C.*

Before proving this, let's see how it implies Proposition 13.1.8.

Proof of Proposition 13.1.8. Fix constants $K \geq 1$ and $\varepsilon \geq 0$ such that the mapping $\phi : \mathcal{X} \to \mathcal{Y}$ is a (K, ε)-quasi-isometry, and let $C = C(K, \varepsilon)$ be as in Proposition 13.2.4. Suppose that T is a geodesic triangle in $\mathcal{X}$ with vertices a, b, c; then since the sides ab, bc, ca are geodesics, the images $\phi(ab), \phi(bc), \phi(ca)$ are (K, ε)-quasi-geodesics in $\mathcal{Y}$, and hence, by Proposition 13.2.4, are within Hausdorff distance C of geodesic segments with endpoints $\phi(a), \phi(b), \phi(c)$, respectively. Since geodesic triangles in $\mathcal{Y}$ are δ-thin, for some $\delta > 0$, it follows that every point of the image $\phi(ab)$ is within distance $2C + \delta$ of $\phi(bc) \cup \phi(ca)$ (and similarly for $\phi(bc)$ and $\phi(ca)$). Consequently, since ϕ is a (K, ε)-quasi-isometry, it follows that every point of ab is within distance $K(2C + \delta) + \varepsilon := \delta'$ of $bc \cup ca$. This proves that $(\mathcal{X}, d_X)$ is a δ'-hyperbolic space. □

For *geodesic* segments, Proposition 13.2.4 is a direct consequence of the thin triangle property. Suppose that β and γ are geodesic segments, both with endpoints x, y. Fix any points z on β and w on γ, and let γ', γ'' be the geodesic segments obtained by partitioning γ at the point w. Then β, γ', γ'' are the sides of a geodesic triangle, and so the point z must lie within distance δ of $\gamma = \gamma' \cup \gamma''$. By symmetry, it follows that w must lie within distance δ of β. Therefore, the Hausdorff distance between β and γ is $\leq \delta$.

The proof of Proposition 13.2.4 for *quasi-geodesic* segments will rely on the following two lemmas, which facilitate comparison of continuous paths with geodesic segments.

Lemma 13.2.5 *If the path* β *is* (K, ε)*-quasi-geodesic then for any points* $x_0, x_1, \cdots, x_n$ *that occur* in order *along* β*,*

$$\sum_{i=0}^{n-1} d(x_i, x_{i+1}) \leq K^2 d(x_0, x_n) + nK\varepsilon + K^2\varepsilon. \tag{13.2.1}$$

Proof. Without loss of generality, assume that x_0 and x_n are the endpoints of β. Let $J \subset \mathbb{R}$ be the interval of parametrization. Since the points $x_0, x_1, \cdots, x_n$ occur in order along β, there exist points $t_0 < t_1 < \cdots < t_n$ in $J = [t_0, t_n]$ such that $\beta(t_i) = x_i$ for each index i. Hence, because the mapping β is (K, ε)-quasi-isometric,

$$\sum_{i=0}^{n-1} d(x_i, x_{i+1}) \leq K \sum_{i=0}^{n-1} (t_{i+1} - t_i) + n\varepsilon = K|J| + n\varepsilon.$$

On the other hand, quasi-isometry also implies that

$$|J| \leq K d(x_0, x_n) + K\varepsilon,$$

and so (13.2.1) follows. □

Lemma 13.2.6 *Let* $(\mathcal{X}, d)$ *be a* δ*-hyperbolic metric space,* β *a continuous path in* $\mathcal{X}$ *with endpoints* x, x'*, and* γ *a geodesic in* $\mathcal{X}$*. Let* $y, y' \in \gamma$ *be geodesic projections of* x, x'*, respectively, onto* γ*. There exist finite sequences of points* $(x_i)_{0 \leq i \leq m}$ *and* $(y_i)_{0 \leq i \leq m}$ *occurring* in order *along* β *and* γ*, respectively, such that*

(a) $x_0 = x$ *and* $x_m = x'$*;*
(b) $y_0 = y$ *and* $y_m = y'$*;*
(c) $8\delta < d(y_i, y_{i+1}) < 16\delta$ *for* $0 \leq i \leq m-2$ *and* $d(y_{m-1}, y_m) < 16\delta$*; and*
(d) for each index i *the point* y_i *is a geodesic projection of* x_i *onto* γ*.*

Proof. The points x_i, y_i are chosen inductively, beginning with $x_0 = x$ and $y_0 = y$. Assume that x_i, y_i have been selected for $i \leq n$ in such a way that the requirements (c)-(d) are satisfied, and so that the points occur in order along β and γ. If $d(y_n, y') < 16\delta$, set $m = n + 1$ and

$$x_m = x' \quad \text{and} \quad y_m = y';$$

the construction is then finished. Otherwise, consider the segment γ' of γ that starts at y_n and ends at y', and let α be the path that first traverses a geodesic segment from y_n to x_n, then the sub-path β' of β from x_n to the terminal endpoint x', and then a geodesic segment from x' to y'. By Exercise 13.1.5 (c), there exists

a point x_{n+1} on this latter path with geodesic projection y_{n+1} onto the geodesic segment γ' such that $8\delta < d(y_n, y_{n+1}) < 16\delta$. By Exercise 13.1.5(b) (specifically, inequality (13.1.5)), the point x_{n+1} cannot lie on another the geodesic segment from y_n to x_n or the geodesic segment from x' to y', so it must lie on β. By construction, the points $(x_i)_{0\le i\le n+1}$ and $(y_i)_{0\le i\le n+1}$ occur in order along β and γ, respectively and conditions (c)–(d) are satisfied. The induction must eventually end, because the geodesic sub-segment of γ from y to y' has finite length. □

Proof of Proposition 13.2.4. Let β be a (K, ε)-quasi-geodesic segment with endpoints x, x', where $K > 1$ and $\varepsilon \ge 0$, and let $\gamma = xx'$ be a geodesic segment with the same endpoints. Without loss of generality (cf. Exercise 13.2.3), assume that β is continuous.

Suppose that for some constant $1 < C < \infty$ the path β is not contained in the $4K^3C$-neighborhood of γ. Since $x \mapsto d(x, \gamma)$ is continuous, β must have a sub-segment β', with endpoints x'', x''', such that

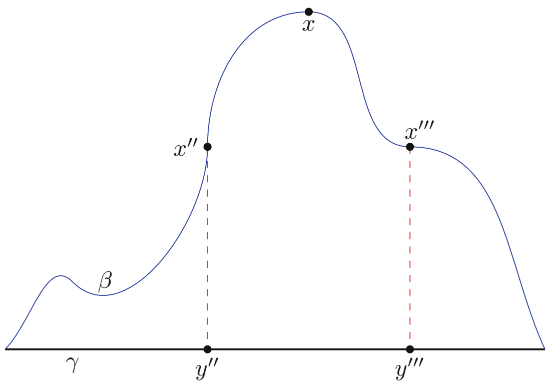
$$d(x'', \gamma) = d(x''', \gamma) = \min_{x\in\beta'} d(x, \gamma) = C \quad \text{and} \tag{13.2.2}$$

$$\max_{x\in\beta'} d(x, \gamma) \ge 4K^3C. \tag{13.2.3}$$

Let y'', y''' be geodesic projections of x'', x''' onto γ (see Exercise 13.1.5). We will exhibit an upper bound on C by (i) using the fact that β is quasi-geodesic to show that if the distance between y'', y''' is $\le 16\delta K^2$ then C must satisfy the inequality (13.2.4) below, and (ii) using δ-hyperbolicity to show that if the distance between y'', y''' is $> 16\delta K^2$ then C must satisfy the inequality (13.2.5).

Case (i): Consider first the case where $d(y'', y''') \le 16\delta K^2$. Denote by β' the sub-segment of β with endpoints x'', x''', and by γ' the sub-segment of γ with endpoints y'', y'''. Since $d(x'', y'') = d(x''', y''') = C$, the triangle inequality implies that $d(x'', x''') \le 2C + 16\delta K^2$. By Lemma 13.2.5, the hypothesis that β is (K, ε)-quasi-geodesic requires that for any point $x \in \beta'$ (see Figure 13.1),

Fig. 13.1 Case (i)



$$d(x'', x) + d(x, x''') \le K^2 d(x'', x''') + 2K\varepsilon + K^2\varepsilon;$$

consequently,

$$\begin{aligned} d(x, \gamma) &\le d(x, x'') + d(x'', y'') \\ &\le K^2 d(x'', x''') + (2K + K^2)\varepsilon + C \\ &\le 2CK^2 + 16\delta K^4 + (2K^2 + K^3)\varepsilon + C. \end{aligned}$$

Therefore, inequality (13.2.3) is possible only if

$$4K^3 C \le 2CK^2 + 16\delta K^4 + (2K^2 + K^3)\varepsilon + C,$$

or equivalently, since $4K^3 - 2K^2 - 1 > 0$,

$$C \le \frac{16\delta K^4 + (2K^2 + K^3)\varepsilon}{4K^3 - 2K^2 - 1}. \tag{13.2.4}$$

Case (ii): Next, consider the case where $d(y'', y''') > 16\delta K^2$. Lemma 13.2.6, when applied to the continuous path β' and the geodesic segment γ', implies that there are corresponding points $\{x_i, y_i\}_{0 \le i \le m}$ on β', γ' such that each y_i is a geodesic projection of x_i onto γ, and such that

$$x_0 = x'', y_0 = y'', x_m = x''', \text{ and } y_m = y'''.$$

Furthermore, Lemma 13.2.6 (c) asserts that the distances between successive points y_i, y_{i+1} are all between 8δ and 16δ (with the possible exception of the final pair y_{m-1}, y_m). By hypothesis $d(y_0, y_m) > 16\delta K^2$, so it follows that $m > 2K^2$. Since the distances $d(y_i, y_{i+1})$ are all at least 8δ, Exercise 13.1.5 (d) implies that any geodesic segment with endpoints x_i, x_{i+1} must pass within distance 2δ of the geodesic segment $y_i y_{i+1}$. (See Figure 13.2.) But by equations (13.2.2) and the fact

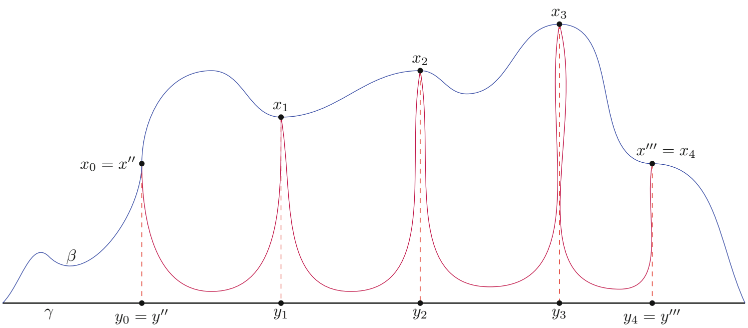


Fig. 13.2 Case (ii)

that each point y_i is a geodesic projection of the corresponding point x_i,

$$d(x_i, y_i) = d(x_i, \gamma) \geq C \quad \text{for each index } i;$$

consequently, the length of any geodesic segment with endpoints x_i and x_{i+1} must be at least $2C - 4\delta$. Thus,

$$\sum_{i=0}^{m-1} d(x_i, x_{i+1}) \geq (m-1)(2C - 4\delta).$$

On the other hand,

$$d(x'', x''') \leq d(x'', y'') + d(y'', y''') + d(y''', x''') \leq 2C + m(16\delta).$$

Hence, by Lemma 13.2.5, since the path β is (K, ε)-quasi-geodesic and $m-1 \geq K^2$,

$$(m-1)(2C - 4\delta) \leq K^2(2C + m(16\delta)) + (m-1)K\varepsilon + K^2\varepsilon \quad \Longrightarrow$$
$$(m - 1 - K^2)(2C) \leq (m-1)(4\delta) + K^2 m(16\delta) + ((m-1)K + K^2)\varepsilon.$$

Since $m - 1 \geq 2K^2$ (and therefore $m - 1 - K^2 \geq (m-1)/2 > 1$), it follows (exercise!) that

$$C \leq 8\delta + 33K^2\delta + (2K + K^2)\varepsilon. \tag{13.2.5}$$

We have proved that if β is not contained in the $4K^3C$-neighborhood of γ then C must satisfy one or the other of the bounds (13.2.4)–(13.2.5), and so there is a constant C_* depending only on K, ε, and δ such that any quasi-geodesic segment β is contained in the C_*-neighborhood of a geodesic segment γ with the same endpoints. Furthermore, any such geodesic segment must contain a finite sequence of points y_i, each at distance $< C_*$ from β, such that every point on γ is within distance 16δ of some y_i. Hence, every point of γ is contained in the $(C_* + 16\delta)$-neighborhood of β, and so the Hausdorff distance between β and γ is bounded by $C_* + 16\delta$. □

The conclusion of Proposition 13.2.4 holds for all quasi-geodesic segments, regardless of their lengths, but not necessarily for infinite quasi-geodesic paths. If however, the ambient hyperbolic space $\mathcal{X}$ is proper then Proposition 13.2.4 extends to infinite quasi-geodesic paths.

Corollary 13.2.7 *If $(\mathcal{X}, d)$ is a proper, δ-hyperbolic metric space, then for any $K \geq 1$ and $\varepsilon \geq 0$ there exists $C = C(K, \varepsilon, \delta) < \infty$ such that any (K, ε)-quasi-geodesic line or ray is within Hausdorff distance C of a geodesic line or ray.*

Proof. Let β be a (K, ε)-quasi-geodesic line in a proper δ-hyperbolic space $(\mathcal{X}, d)$ such that any finite sub-segment is rectifiable, and assume that β is parametrized by

arc length. Set $x = \beta(0)$. By definition,

$$\liminf_{|t|\to\infty} \frac{1}{|t|} d(\beta(t), x) \geq K^{-1},$$

and so in particular $\lim_{n\to\infty} d(\beta(\pm n), x) = \infty$.

For each $n \in \mathbb{N}$ let $\beta([-n, n])$ be the quasi-geodesic segment obtained by restricting the map β to the interval $[-n, n]$. By Proposition 13.2.4, any geodesic segment γ_n connecting the endpoints $\beta(-n)$ and $\beta(+n)$ must lie within Hausdorff distance $\leq C = C(K, \varepsilon, \delta)$ of $\beta([-n, n])$. In particular, each γ_n intersects the closed ball $\mathbb{B}_C(x)$ of radius C centered at x, so there is a parametrization $\gamma_n : J_n \to \mathcal{X}$ of γ_n with $0 \in J_n$ such that $\gamma_n(0) \in \mathbb{B}_C(x)$. Since γ_n has endpoints $\beta(-n)$ and $\beta(+n)$ that are, for large n, distant from x, the intervals J_n must, for large n, contain any fixed interval $[-m, m]$. By definition, each of the mappings γ_n is an isometry, so the restrictions of these mappings to the interval $[-m, m]$ form an *equicontinuous* family, all with images contained in the closed ball $\mathbb{B}_{m+C}(x)$. Now any such ball is compact, by virtue of the Heine-Borel property, so the *Arzela-Ascoli* Theorem (see, e.g., [101], Theorem 47.1) implies that for each $m \in \mathbb{N}$ there is a subsequence n_j such that the restrictions of the mappings γ_{n_j} to the interval $[-m, m]$ converge uniformly. Using Cantor's diagonal trick, we can extract a subsequence n_j such that the mappings γ_{n_j} converge uniformly on every interval $[-m, m]$. The image of the limit mapping is easily seen to be a doubly-infinite geodesic γ, and by construction this lies within Hausdorff distance C of β.

The proof for geodesic rays is virtually identical. □

13.3 The Gromov Boundary of a Hyperbolic Space

Every compact geodesic metric space is trivially δ-hyperbolic for some $\delta < \infty$ (for instance, take δ to be the diameter of the space). Hyperbolicity only has nontrivial consequences for *unbounded* spaces, which have natural compactifications in which geodesic rays converge.

Assumption 13.3.1 *Assume in this section that $(\mathcal{X}, d)$ is a* proper*, δ-hyperbolic metric space, and let $z \in \mathcal{X}$ be a fixed reference point.*

Definition 13.3.2 The *Gromov product* of two points $x, y \in \mathcal{X}$ with respect to the reference point $z \in \mathcal{X}$ is defined by

$$(x|y)_z = \frac{1}{2}(d(z, x) + d(z, y) - d(x, y)). \tag{13.3.1}$$

The definition (13.3.1) is valid in any metric space, but is only of real interest for hyperbolic spaces. If the ambient metric space is δ-hyperbolic, then to within an

error of size 4δ, the Gromov product $(x|y)_z$ is the distance from z to any geodesic segment connecting x to y.

Exercise 13.3.3 Prove this; in particular, show that for any geodesic segment β with endpoints x and y,

$$d(z, \beta) - 4\delta \leq (x|y)_z \leq d(z, \beta). \tag{13.3.2}$$

HINT: See Exercise 13.1.5 (b).

To within an error of size 4δ, the Gromov product $(x|y)_z$ is the distance that geodesic segments γ_x, γ_y from z to x, y remain within distance δ. (See Exercise 13.3.16 below.) By the triangle inequality,

$$|(x|y)_z - (x|y)_{z'}| \leq d(z, z') \quad \text{for all } x, y, z, z' \in \mathcal{X}. \tag{13.3.3}$$

Definition 13.3.4 A sequence $(x_n)_{n\in\mathbb{N}}$ of points in $(\mathcal{X}, d)$ *converges to infinity* if for some (every) reference point z,

$$\liminf_{i,j\to\infty} (x_i|x_j)_z = \infty. \tag{13.3.4}$$

The inequality (13.3.3) implies that the property (13.3.4) does not depend on the reference point z. Condition (13.3.4) is more stringent than the condition $\lim_{n\to\infty} d(x_n, z) = \infty$, which is the usual meaning of the term *converges to infinity*; however, Definition 13.3.4 is standard terminology in hyperbolic geometry.

Definition 13.3.5 Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be two sequences in $\mathcal{X}$, both of which converge to infinity. The sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are *equivalent at infinity* (or, more briefly, *equivalent*) if for some (every) reference point z,

$$\liminf_{i,j\to\infty} (x_i|y_j)_z = \infty. \tag{13.3.5}$$

The *Gromov boundary* $\partial\mathcal{X}$ is the set of equivalence classes of sequences that converge to infinity.

Exercise 13.3.6 (Rough Ultrametric Property of the Gromov Product) Justification of Definition 13.3.5 requires proof that *equivalence* of sequences is really an equivalence relation, and in particular is *transitive*.

(A) Prove the following *rough ultrametric property*: for any $x, y, w \in \mathcal{X}$,

$$(x|w)_z \geq \min((x|y)_z, (w|y)_z) - 4\delta. \tag{13.3.6}$$

(B) Deduce from this that equivalence of convergent sequences is transitive.

Exercise 13.3.7 Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{X}$ such that $(x_n)_{n\in\mathbb{N}}$ converges to infinity. Show that if the sequence of real numbers $d(x_n, y_n)$ is

bounded, then the sequence $(y_n)_{n\in\mathbb{N}}$ converges to infinity and is equivalent to the sequence $(x_n)_{n\in\mathbb{N}}$.

The Gromov product extends to the space $\mathcal{X} \cup \partial\mathcal{X}$ as follows: (1) for $x \in \mathcal{X}$ and $\xi \in \partial\mathcal{X}$,

$$(x|\xi)_z := \sup \liminf_{i\to\infty} (x|x_i)_z \tag{13.3.7}$$

where the supremum is taken over all sequences $(x_i)_{i\in\mathbb{N}}$ in the equivalence class ξ, and (2) for $\xi, \zeta \in \partial\mathcal{X}$,

$$(\xi|\zeta)_z := \sup \liminf_{i,j\to\infty} (x_i|y_j), \tag{13.3.8}$$

where the sup is over all sequences $(x_i)_{i\in\mathbb{N}}$ and $(y_j)_{j\in\mathbb{N}}$ in the equivalence classes ξ, ζ, respectively.

Exercise 13.3.8

(A) Show that $(\xi|\zeta)_z < \infty$ unless $\xi = \zeta$.
(B) Show that if $\xi, \zeta \in \partial\mathcal{X}$ then for any two sequences $(x_i)_{i\in\mathbb{N}}$ and $(y_j)_{j\in\mathbb{N}}$ representating ξ and ζ, respectively, we have

$$(\xi|\zeta)_z - 2\delta \le \liminf_{i,j\to\infty} (x_i|y_j)_z \le (\xi|\zeta)_z.$$

(C) Show that the rough ultrametric property (13.3.6) extends to $\mathcal{X} \cup \partial\mathcal{X}$ (perhaps with the error 4δ increased to 6δ or 8δ).

The Gromov product can be used to define a topology on $\mathcal{X} \cup \partial\mathcal{X}$, called the *Gromov topology*, whose restriction to the space $\mathcal{X}$ is the topology induced by the metric d. Closed neighborhoods of points $x \in \mathcal{X}$ are just balls $\mathbb{B}_r(x) := \left\{x' \in \mathcal{X} \,:\, d(x,x') \le r\right\}$ of positive radius r centered at x; for points $\xi \in \partial\mathcal{X}$ and scalars $r \ge 0$, neighborhoods $V(\xi; r)$ are defined by

$$V(\xi; r) := \{a \in \mathcal{X} \cup \partial\mathcal{X} \,:\, (\xi|a)_z \ge r\}. \tag{13.3.9}$$

A sequence of points $\xi^{(n)}$ in $\partial\mathcal{X}$ converges to $\xi \in \partial\mathcal{X}$ in the Gromov topology if and only if

$$\lim_{n\to\infty} (\xi^{(n)}|\xi)_z = \infty. \tag{13.3.10}$$

Exercise 13.3.9 Show that the sets $V(\xi; r)$ and $\mathbb{B}_r(x)$ form a neighborhood basis for a Hausdorff topology on $\mathcal{X} \cup \partial\mathcal{X}$.

HINT: To prove that these sets constitute a neighborhood basis, you must show that for any $\zeta \in V(\xi_1; r_1) \cap V(\xi_2; r_2)$ there exists r_3 such that $V(\zeta; r_3) \subset V(\xi_1; r_1) \cap V(\xi_2; r_2)$. For this, use the rough ultrametric property of the Gromov product. To

prove that the topology is Hausdorff, it suffices to show that for any two distinct points $\xi, \zeta \in \partial\mathcal{X}$ there exists $r < \infty$ such that $V(\xi; r) \cap V(\zeta; r) = \emptyset$.

Proposition 13.3.10 *The Gromov topology on $\mathcal{X} \cup \partial\mathcal{X}$ is metrizable: in particular, for any sufficiently small $\varepsilon > 0$ there is a metric d_ε such that for some $C > 0$ depending on ε,*

$$Ce^{-\varepsilon(\xi|\zeta)_z} \leq d_\varepsilon(a, b) \leq e^{-\varepsilon(\xi|\zeta)_z} \quad \textit{for all } \xi, \zeta \in \mathcal{X} \cup \partial\mathcal{X}. \tag{13.3.11}$$

Proof Sketch. For any two points ξ, ζ define

$$d_\varepsilon(\xi, \zeta) = \inf \sum_{i=1}^{n} e^{-\varepsilon(\eta_{i-1}|\eta_i)_z}$$

where the inf is over all finite sequences $(\eta_i)_{0\leq i\leq n}$ with initial term $\eta_0 = \xi$ and final term $\eta_n = \zeta$. By definition, the upper inequality in (13.3.11) holds, so it suffices to show that if $\varepsilon > 0$ is sufficiently small then the lower inequality holds. This can be done by proving inductively that for any $n \in \mathbb{N}$ and any finite sequence $(\eta_i)_{0\leq i\leq n}$ connecting ξ and ζ,

$$C \exp\{-\varepsilon(\xi|\zeta)_z\} \leq \sum_{i=1}^{n} \exp\{-\varepsilon(\eta_{i-1}|\eta_i)_z\}$$

for some $C > 0$ depending only on ε and the hyperbolicity constant δ. □

Exercise 13.3.11 † Finish the argument.

HINT: For the induction step, use the rough ultrametric property of Exercise 13.3.6. See Ghys and de la Harpe [50], Chapter 7, Section 3 for a complete solution.

Working with the Gromov topology is facilitated by the fact that points of $\partial\mathcal{X}$ have "geodesic representatives". This is explained by the next two propositions.

Proposition 13.3.12 *Let β be a quasi-geodesic ray in $\mathcal{X}$.*

(A) If $(x_n)_{n\in\mathbb{N}}$ is any sequence of points on β such that $\lim_{n\to\infty} d(x_n, z) = \infty$, then the sequence $(x_n)_{n\in\mathbb{N}}$ converges in the Gromov topology to a point $\xi \in \partial\mathcal{X}$.
(B) If β' is any quasi-geodesic ray whose Hausdorff distance from β is finite, then for any sequence of points $(x'_n)_{n\in\mathbb{N}}$ on β' such that $\lim_{n\to\infty} d(x'_n, z) = \infty$, the sequences $(x_n)_{n\in\mathbb{N}}$ and $(x'_n)_{n\in\mathbb{N}}$ are equivalent, and therefore converge to the same point in the Gromov boundary.

Proof. By Corollary 13.2.7, there is a geodesic ray γ whose Hausdorff distance C from β is finite. For each $n \in \mathbb{N}$, let y_n be a geodesic projection of x_n onto γ; then $d(x_n, y_n) \leq C$, and so by Exercise 13.3.7, the sequence $(x_n)_{n\in\mathbb{N}}$ converges to infinity if and only if the sequence $(y_n)_{n\in\mathbb{N}}$ does. Thus, in proving the first assertion it suffices to consider the case where $\beta = \gamma$ is a geodesic ray. Moreover, we can

assume without loss of generality that the reference point z is $z = y_0$ (by (13.3.3)), and that the points y_n occur in order along γ (by taking subsequences). But in this case, if $n > m$ we have

$$(y_n|y_m)_z = d(z, y_m) \to \infty.$$

□

Exercise 13.3.13 Prove the second assertion.

HINT: First show that without loss of generality we may take β' to be a *geodesic* ray.

Proposition 13.3.12 implies that for any quasi-geodesic ray β there is a *unique* point $\xi \in \partial\mathcal{X}$ to which points on β can converge. Henceforth, we will say that the quasi-geodesic ray β *represents* ξ, or alternatively that β *converges to* ξ.

Proposition 13.3.14 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of points in $\mathcal{X}$ that converges to infinity. For any point $y_0 \in \mathcal{X}$ there is a geodesic ray γ with initial point y_0 such that any sequence of points $(y_n)_{n\in\mathbb{N}}$ on γ satisfying $d(y_n, y_0) \to \infty$ converges to infinity and is equivalent to the sequence $(x_n)_{n\in\mathbb{N}}$.*

Proof. For each $n \in \mathbb{N}$, let γ_n be a geodesic segment connecting y_0 to x_n, with arc length parametrization $\gamma_n : [0, L_n] \to \mathcal{X}$. Since $d(z, x_n) \to \infty$, we must also have $L_n \to \infty$. The Arzela-Ascoli theorem implies that the sequence $(\gamma_n)_{n\in\mathbb{N}}$ is equicontinuous, and hence, since the ambient metric space $\mathcal{X}$ is proper, has a subsequence that converges uniformly on compact intervals of $[0, \infty)$. The limit function $\gamma : [0, \infty)$ defines a geodesic ray with initial point y_0.

For any point y on γ, the geodesic segments γ_n come increasingly close to y as $n \to \infty$, that is, $\lim_{n\to\infty} d(y, \gamma_n) = 0$. Hence, by definition of the Gromov product and inequality (13.3.3)

$$\lim_{n\to\infty} (y|x_n)_{y_0} = d(y_0, y) \quad \Longrightarrow$$

$$\liminf_{n\to\infty} (y|x_n)_z \geq d(y_0, y) - d(z, y_0).$$

It follows that for any sequence of points y_m on γ such that $d(y_m, y_0) \to \infty$,

$$\lim_{m\to\infty} \liminf_{n\to\infty} (y_m|x_n)_z = \infty,$$

and so the sequences $(x_n)_{n\geq 1}$ and $(y_m)_{m\geq 1}$ are equivalent at infinity. □

There can be more than one geodesic ray converging to a point $\xi \in \partial\mathcal{X}$; in fact, there can be infinitely many, even when the hyperbolic space is the Cayley graph of a hyperbolic group. (Example: $\Gamma = \mathbb{Z} \oplus \mathbb{Z}_2$; for this group, with the standard generators, the Cayley graph is an infinite "ladder", and its Gromov boundary has

exactly two points, one at each "end" of the ladder.) Nevertheless, the following is true.

Proposition 13.3.15 *Two quasi-geodesic rays both converging to a point $\xi \in \partial\mathcal{X}$ must be at finite Hausdorff distance from one another.*

Proof. Without loss of generality (by Corollary 13.2.7) it suffices to prove this for *geodesic* rays. Let γ_1, γ_2 be geodesic rays both converging to $\xi \in \partial\mathcal{X}$. Let $x_n = \gamma_1(n)$ and $y_n = \gamma_2(n)$ be the points on γ_1, γ_2 at distance n from the initial points of the rays. By Proposition 13.3.14, these sequences are equivalent and both converge to ξ, and so $(x_n|y_n)_z \to \infty$. Now $(x_n|y_n)_z$ is, to within an error of size 4δ, the distance from the reference point z to any geodesic segment β_n with endpoints x_n and y_n; consequently, by the triangle inequality, if α is a geodesic segment with endpoints $\gamma_1(0)$ and $\gamma_2(0)$ then

$$\lim_{n\to\infty} d(\alpha, \beta_n) = \infty.$$

Let Q_n be the geodesic quadrilateral with sides α, β_n, $\gamma_1[0, n]$, and $\gamma_2[0, n]$. By Exercise 13.1.4, any point on the segment $\gamma_1[0, n]$ (respectively $\gamma_2[0, n]$) not within distance 4δ of either α or β_n must be within distance 2δ of $\gamma_2[0, n]$ (respectively $\gamma_1[0, n]$). Letting $n \to \infty$, we see that every point on γ_1 (respectively, γ_2) not within distance 4δ of α must be within distance 2δ of γ_2 (respectively, γ_1). Letting $n \to \infty$, one sees that, except for their initial segments, the rays γ_1 and γ_2 are within Hausdorff distance 2δ. □

Exercise 13.3.16 Let α and β be geodesic rays or segments, both with initial point z, and with endpoints $\xi, \zeta \in \mathcal{X} \cup \partial\mathcal{X}$ respectively. Show that if $(\xi|\zeta)_z = C$, then the geodesics "separate" at distance $\approx C$, in the following sense:

$$\begin{aligned} d(\alpha(s), \beta) &\geq s - C - 2\delta - 1 \quad \text{for all } s \geq C + 4\delta; \\ d(\beta(s), \alpha) &\geq s - C - 2\delta - 1 \quad \text{for all } s \geq C + 4\delta; \quad \text{and} \\ d(\alpha(t), \beta(t)) &\leq 2\delta + 1 \quad \text{for all } t \leq C - 4\delta. \end{aligned} \tag{13.3.12}$$

Thus, in particular, if geodesic rays α and β both converge to the same boundary point then $d(\alpha(t), \beta(t)) \leq 2\delta + 1$ for all $t \geq 0$.

HINT: The thin triangle property guarantees that for $t \leq C - 4\delta$ we have $d(\alpha(t), \beta) \leq \delta$ and $d(\alpha, \beta(t)) \leq \delta$. Show that if the third inequality in (13.3.12) did not hold then at least one of the paths α, β would not be geodesic.

Propositions 13.3.12 and 13.3.14 imply that for a proper, hyperbolic metric space $(\mathcal{X}, d)$, every equivalence class in the Gromov boundary $\partial\mathcal{X}$ has a geodesic representative, and conversely, every quasi-geodesic ray determines an element of $\partial\mathcal{X}$. The Gromov topology has the following useful characterization in terms of geodesic representatives.

Proposition 13.3.17 *A sequence $(\xi_n)_{n\in\mathbb{N}} \subset \partial\mathcal{X}$ converges to $\xi \in \partial\mathcal{X}$ in the Gromov topology if and only if every subsequence of $(\xi_n)_{n\in\mathbb{N}}$ has a subsequence $(\xi_m)_{m\in\mathbb{N}}$ with geodesic representatives $\gamma_m : [0,\infty) \to \mathcal{X}$ such that γ_m converges as $m \to \infty$ to a geodesic representative γ of ξ uniformly on bounded intervals.*

Proof. Suppose that the elements $\xi_n \in \partial\mathcal{X}$ have geodesic representatives $\gamma_n : [0,\infty) \to \mathcal{X}$, all with initial point z, such that γ_n converge uniformly on bounded intervals to a geodesic representative γ of $\xi \in \partial\mathcal{X}$. Fix $R > 0$ large. Since each γ_n is geodesic,

$$d(z, \gamma_n(t)) = t \geq 2R \quad \text{for all } t \geq 2R$$

Since $\gamma_n \to \gamma$ uniformly on the interval $[0, 2R]$, it follows that for all sufficiently large n the geodesics γ_n and γ exit the ball $\mathbb{B}_R(z)$ at points y' and y such that $d(y', y) < 1$. Hence, by definition of the Gromov product, for any points $w \in \gamma[2R, \infty)$ and $w' \in \gamma_n[2R, \infty)$,

$$\begin{aligned}(w|w')_z &= \frac{1}{2}\big(d(z,w) + d(z,w') - d(w,w')\big)\\ &\geq \frac{1}{2}\big(d(z,w) + d(z,w') - d(y,w) - d(y',w') - d(y,y')\big)\\ &= \frac{1}{2}\big(d(z,w) + d(z,w') - d(y,y')\big) \geq R - \frac{1}{2}.\end{aligned}$$

Since R can be taken arbitrarily large, this proves that $\lim_{n\to\infty}(\xi_n|\xi)_z = \infty$.

Next, suppose that $\xi_n \to \xi$ in the Gromov topology. Let $\gamma_n : [0,\infty) \to \mathcal{X}$ be a geodesic representative of ξ_n with initial point $\gamma_n(0) = z$. Since the metric space $(\mathcal{X}, d)$ is proper, the Arzela-Ascoli theorem implies that any subsequence of $(\gamma_n)_{n\in\mathbb{N}}$ has a subsequence $(\gamma_m)_{m\in\mathbb{N}}$ that converges uniformly on bounded intervals. The limit is a geodesic ray γ with initial point z; by Proposition 13.3.12, this geodesic ray represents some $\zeta \in \partial\mathcal{X}$. By the preceding paragraph, we must have $\xi_n \to \zeta$ in the Gromov topology, and therefore $\zeta = \xi$. □

Corollary 13.3.18 *The space $\mathcal{X} \cup \partial\mathcal{X}$ is compact in the Gromov topology.*

Proof. Since the Gromov topology is metrizable, compactness is equivalent to sequential compactness, so it suffices to establish the latter. First, let $(\xi^{(n)})_{n\in\mathbb{N}}$ be a sequence of points in the boundary $\partial\mathcal{X}$. Each $\xi^{(n)}$ has a geodesic representative γ_n with initial point $\gamma_n(0) = z$. Since $\mathcal{X}$ has the Heine-Borel property, Arzela-Ascoli implies that the sequence $(\gamma_n)_{n\in\mathbb{N}}$ has a subsequence that converges uniformly on bounded intervals. By Proposition 13.3.17, the corresponding subsequence of $(\xi^{(n)})_{n\in\mathbb{N}}$ converges in the Gromov topology.

Next, let $(x_n)_{n\in\mathbb{N}}$ be a sequence of points in $\mathcal{X}$. If this sequence is bounded, then it must have a convergent subsequence, because by hypothesis bounded subsets of $\mathcal{X}$ are pre-compact. If the sequence is unbounded, then we may assume, by passing to a subsequence if necessary, that $d(x_n, z) \to \infty$. Let γ_n be a geodesic segment

from z to x_n. Since $d(x_n, z) \to \infty$, for any $R < \infty$ all but finitely many of the geodesic segments γ_n must exit the ball $\mathbb{B}_R(z)$. Thus, by Arzela-Ascoli, there is a subsequence $(\gamma_m)_{m\in\mathbb{N}}$ that converges uniformly on bounded intervals to a geodesic ray γ. By the same argument as in the proof of Proposition 13.3.17, we conclude that $x_n \to \xi$, where $\xi \in \partial\mathcal{X}$ is the equivalence class represented by γ. □

Corollary 13.3.19 *Any quasi-isometric mapping $T : \mathcal{X} \to \mathcal{Y}$ of proper, hyperbolic metric spaces $(\mathcal{X}, d_X)$ and $(\mathcal{Y}, d_Y)$ extends to a continuous mapping $\hat{T} : \partial\mathcal{X} \to \partial\mathcal{Y}$ relative to the Gromov topologies. The extension has the following property: for any sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ that converges to $\xi \in \partial\mathcal{X}$, the sequence $(Tx_n)_{n\in\mathbb{N}}$ converges to $\hat{T}\xi \in \partial\mathcal{Y}$.*

Proof. To prove that the extension $\hat{T}$ is well-defined, we must show that if $(x_n)_{n\in\mathbb{N}}$ and $(x'_n)_{n\in\mathbb{N}}$ are two sequences in $\mathcal{X}$ that both converge to $\xi \in \partial\mathcal{X}$, then the sequences $(Tx_n)_{n\in\mathbb{N}}$ and $(Tx'_n)_{n\in\mathbb{N}}$ both converge to the same point $\zeta \in \partial\mathcal{Y}$. If $(x_n)_{n\in\mathbb{N}}$ and $(x'_n)_{n\in\mathbb{N}}$ converge to the same boundary point, then they are equivalent sequences, and so $(x_n|x'_n)_z \to \infty$. The Gromov product $(x_n|x'_n)_z$ is, up to an error of size 2δ, the distance from z to a geodesic segment α_n from x_n to x'_n; thus, $\lim_{n\to\infty} d_X(z, \alpha_n) = \infty$. The mapping T distorts distances by a factor of at most $K^{\pm 1}$, so it follows that $\lim_{n\to\infty} d_Y(Tz, T\alpha_n) = \infty$. Since the mapping T is (K, ε)-quasi-geodesic, the path $T\alpha_n$ is quasi-geodesic, and so by Proposition 13.2.4 it lies within Hausdorff distance $C = C(K, \varepsilon, \delta)$ of a *geodesic* segment β_n from Tx_n to Tx'_n. Thus, $\lim_{n\to\infty} d(Tz, \beta_n) = \infty$, and so $\lim_{n\to\infty}(Tx_n|Tx'_n)_{Tz} = \infty$. This proves that the sequences $(Tx_n)_{n\in\mathbb{N}}$ and $(Tx'_n)_{n\in\mathbb{N}}$ are equivalent, and therefore converge to the same boundary point $\hat{T}\xi$.

A similar argument (Exercise!) proves that the extension $\hat{T} : \partial\mathcal{X} \to \partial\mathcal{Y}$ is continuous. □

It is noteworthy that Corollary 13.3.19 does not require that the mapping T be continuous. There are, in fact, interesting examples of quasi-isometric mappings that are *not* continuous: for instance, if $\mathbb{A}$ and $\mathbb{A}'$ are two finite, symmetric generating sets of a hyperbolic group Γ, the identity map $\mathrm{id} : \Gamma \to \Gamma$ extends to a quasi-isometric mapping $T : G_{\Gamma;\mathbb{A}} \to G_{\Gamma;\mathbb{A}'}$ that is, in general, not continuous. Corollary 13.3.19 ensures that any two such Cayley graphs have the same Gromov boundary.

Proposition 13.3.12 implies that for any geodesic line γ the limits $\xi_\pm = \lim_{t\to\pm\infty} \gamma(t)$ exist, and Proposition 13.3.15 together with the thin triangle property implies that these are distinct. Call these points $\xi_\pm$ the *endpoints* of γ.

Proposition 13.3.20 *Any two distinct points $\xi, \zeta \in \partial\mathcal{X}$ are the endpoints of a geodesic line in $\mathcal{X}$.*

Proof. Proposition 13.3.12 implies that for any two distinct points $\xi, \zeta \in \partial\mathcal{X}$ there are geodesic rays α, β, both with initial point z, that converge to ξ, ζ, respectively. Let $(\xi|\zeta)_z = C$; then $C < \infty$, since the limit points ξ, ζ are distinct. Thus, the rays α and β must separate after distance $\approx C$: in particular,

$$d(\alpha(t), \beta(t)) > 2\delta \quad \text{for all } t > C + 2\delta.$$

Consequently, by the thin triangle property, any geodesic segments γ_n connecting $\alpha(n)$ to $\beta(n)$ must all pass within distance 4δ of the separation point(s), and hence within distance $C + 6\delta$ of the reference point z. Assume that the geodesics γ_n are parametrized so that for each n the point $\gamma_n(0)$ is a closest point on γ_n to z, so that $d(\gamma_n(0), z) \leq C + 6\delta$. Then by Arzela-Ascoli, some subsequence of $(\gamma_n)_{n\in\mathbb{N}}$ converges uniformly on bounded intervals to a bi-infinite geodesic γ. By construction, this geodesic line satisfies

$$\lim_{t\to-\infty} \gamma(-t) = \xi \quad \text{and} \quad \lim_{t\to\infty} \gamma(t) = \zeta \tag{13.3.13}$$

in the Gromov topology. □

Proposition 13.3.21 *Let $(\xi_n)_{n\in\mathbb{N}}$ and $(\zeta_n)_{n\in\mathbb{N}}$ be sequences of points in the Gromov boundary $\partial\mathcal{X}$ such that $\xi_n \neq \zeta_n$, and for each $n \in \mathbb{N}$ let γ_n be a geodesic line with endpoints ξ_n and ζ_n. Assume that γ_n is parametrized so that $\gamma_n(0)$ is a closest point on γ_n to the reference point z, that is, $d(z, \gamma_n(0)) = \min_{t\in\mathbb{R}} d(z, \gamma_n(t))$. If*

$$\lim_{n\to\infty} \xi_n = \xi \quad \textit{and} \quad \lim_{n\to\infty} \zeta_n = \zeta \tag{13.3.14}$$

in the Gromov topology, where $\xi \neq \zeta$, then there is a subsequence $(\gamma_k)_{k\in\mathbb{N}}$ that converges uniformly on compact intervals to a geodesic line γ with endpoints ξ and ζ.

Proof. It suffices to show that for some subsequence $(\gamma_m)_{m\in\mathbb{N}}$ the points $\gamma_m(0)$ all lie within a bounded distance of z, because if this is the case then by Arzela-Ascoli the sequence γ_m will have a convergent subsequence, and the limit of any such subsequence will be a geodesic line connecting ξ and ζ.

By Proposition 13.3.14, there are geodesic rays α_n^+, α_n^- and α^+, α_-, all with initial point z, that converge in the Gromov topology to ξ_n, ζ_n and ξ, ζ, respectively. Since $\xi_n \to \xi$ and $\zeta_n \to \zeta$, Proposition 13.3.17 implies that, by passing to a subsequence if necessary, we may assume that

$$\lim_{n\to\infty} \alpha_n^{\pm}(t) = \alpha^{\pm}(t) \tag{13.3.15}$$

uniformly for t in any bounded interval $[0, T]$. Let α_n and α be the bi-infinite paths defined by

$$\begin{aligned} \alpha_n(t) = \alpha_n^+(+t) \quad &\text{and} \quad \alpha(t) = \alpha^+(+t) \quad \text{for } t \geq 0, \\ \alpha_n(t) = \alpha_n^-(-t) \quad &\text{and} \quad \alpha(t) = \alpha^-(-t) \quad \text{for } t \leq 0. \end{aligned}$$

We will show that for some $K \geq 2$ and $\varepsilon \geq 1$ the paths α_m are, at least for all sufficiently large m, (K, ε)-quasi-geodesic. It will then follow, by Corollary 13.2.7, that for some $C = C(K, \varepsilon) < \infty$ and all m sufficiently large, the geodesic γ_m is within Hausdorff distance C of α_m, and so $d(\gamma_m(0), z) \leq C$.

Each path α_n is obtained by splicing together two geodesic rays α_n^+, α_n^-, so to show that the path α_n is (K, ε)-quasi-geodesic, it suffices to show that

$$d(\alpha_n^+(s), \alpha_n^-(t)) \geq K^{-1}(s+t) - \varepsilon \quad \text{for all } s, t \geq 0. \tag{13.3.16}$$

For all sufficiently large s, t we must have

$$d(\alpha^+(t), \alpha^-(s)) \geq \frac{1}{2}(s+t),$$

because otherwise there would be sequences $s_n, t_n \to \infty$ such that $(\alpha^+(s_n)|\alpha^-(t_n))_z = \infty$, contradicting the fact that the geodesic rays α^+ and α^- converge to distinct points of the Gromov boundary. Hence, for some constants $K' \geq 2, \varepsilon' \geq 1$ the path α is (K', ε')-quasi-geodesic. Now the geodesic rays $\alpha_n^\pm$ converge to $\alpha^\pm$ uniformly on bounded intervals, so for any $T < \infty$ there exists $n_T \in \mathbb{N}$ such that if $n \geq n_T$ then (13.3.16) holds for all $s, t \leq T$, with $K = 2K'$ and $\varepsilon = 2\varepsilon'$. Thus, if (13.3.16) fails globally for infinitely many $n \in \mathbb{N}$ then there must exist subsequences $s_k, t_k, n_k \to \infty$ such that

$$d(\alpha_{n_k}^+(s_k), \alpha_{n_k}^-(t_k)) \geq K^{-1}(s_k + t_k) - \varepsilon; \tag{13.3.17}$$

since $K^{-1} \leq \frac{1}{2}$, this implies that

$$\lim_{k\to\infty} (\alpha_{n_k}^+(s_k)|\alpha_{n_k}^-(t_k))_z = \infty, \tag{13.3.18}$$

and so the sequences $(\alpha_{n_k}^+(s_k))_{k\in\mathbb{N}}$ and $(\alpha_{n_k}^-(t_k))_{k\in\mathbb{N}}$ are equivalent. But by construction, $(\alpha_{n_k}^+(s_k))_{k\in\mathbb{N}}$ converges to the boundary point ξ and $(\alpha_{n_k}^-(t_k))_{k\in\mathbb{N}}$ converges to $\zeta \neq \xi$, so the two sequences cannot be equivalent. This proves that (13.3.17) is impossible, so for all large n the path α_n must be (K, ε)-quasi-geodesic. □

13.4 Boundary Action of a Hyperbolic Group

Assumption 13.4.1 *Assume for the remainder of this chapter that* Γ *is a hyperbolic group with* δ*-hyperbolic Cayley graph* $G_{\Gamma;\mathbb{A}}$ *and word metric* d*. Denote by* $\partial\Gamma$ *the Gromov boundary.*

Any two Cayley graphs of Γ are quasi-isometric, by a mapping whose restriction to Γ is the identity, so their Gromov boundaries are homeomorphic. This justifies the notation $\partial\Gamma$ for the boundary. Each group element $g \in \Gamma$ acts as an isometry on $G_{\Gamma;\mathbb{A}}$, and hence, by Corollary 13.3.19, induces a homeomorphism $\zeta \mapsto g \cdot \zeta$ of the boundary $\partial\Gamma$. A crucial feature of this group action is the following contractivity property.

Proposition 13.4.2 *For any sequence* $(g_n)_{n\in\mathbb{N}}$ *of groups elements such that* $\lim_{n\to\infty} g_n = \xi \in \partial\Gamma$ *in the Gromov topology there is a subsequence* $(g_m)_{m\in\mathbb{N}}$ *and a point* $\zeta \in \partial\Gamma \setminus \{\xi\}$ *such that*

$$\lim_{m\to\infty} g_m \cdot \eta = \xi \quad \textit{for every } \eta \in \partial\Gamma \setminus \{\zeta\}. \tag{13.4.1}$$

Remark 13.4.3 If the group elements g_n converge to a point $\xi \in \partial\Gamma$, then for any $h \in \Gamma$ the sequence $(g_n h)_{n\in\mathbb{N}}$ also converges to ξ, because the distances $d(g_n h, g_n) = d(h, 1)$ are uniformly bounded: see Exercise 13.3.7. In fact, the convergence $\lim_{n\to\infty} g_n h = \xi$ holds *uniformly* for h in any finite subset of Γ.

Proof of Proposition 13.4.2. If the mappings $\eta \mapsto g_n \cdot \eta$ converge to the constant mapping $\eta \mapsto \xi$ then there is nothing to prove, so assume that there is a point $\zeta \in \partial\Gamma$ such that the sequence $(g_n \cdot \zeta)_{n\in\mathbb{N}}$ does not converge to ξ. Since $\partial\Gamma$ is compact, by Corollary 13.3.18, there exists a point $\zeta' \neq \xi$ in $\partial\Gamma$ and a subsequence $(g_m)_{m\in\mathbb{N}}$ such that $\lim_{m\to\infty} g_m \cdot \zeta = \zeta'$. We will argue that the convergence (13.4.1) must hold for every $\eta \neq \zeta$.

Suppose otherwise; then there exist points $\zeta'' \neq \zeta$ and $\zeta''' \neq \xi$ and a subsequence $(g_k)_{k\in\mathbb{N}}$ of $(g_m)_{m\in\mathbb{N}}$ such that $\lim_{k\to\infty} g_k \cdot \zeta'' = \zeta'''$. Since $\zeta \neq \zeta''$, there is a geodesic line γ connecting ζ and ζ'' (cf. Proposition 13.3.20); and since each group element g_k acts on the Cayley graph as an isometry, for each $k \in \mathbb{N}$ the image $\gamma_k := g_k \cdot \gamma$ is a geodesic line connecting the boundary points $g_k \cdot \zeta$ and $g_k \cdot \zeta''$. This geodesic line contains the point $g_k \cdot \gamma(0)$, and by Remark 13.4.3,

$$\lim_{k\to\infty} g_k \cdot \gamma(0) = \xi \tag{13.4.2}$$

in the Gromov topology. We will show that this contradicts the hypothesis that the geodesic lines $\gamma_k := g_k \cdot \gamma$ have endpoints on $\partial\Gamma$ that converge to points $\zeta' \neq \xi$ and $\zeta''' \neq \xi$.

By Proposition 13.3.21, we may assume (by passing to a subsequence if necessary) that, after suitable reparametrizations, the geodesics γ_k converge, uniformly on bounded intervals, to a geodesic line β connecting the boundary points ζ' and ζ'''. Thus, since the ambient metric space is a Cayley graph, for any $R \in \mathbb{N}$ there exists $k_R \in \mathbb{N}$ such that

$$\gamma_k(t) = \beta(t) \quad \text{for all } |t| \leq R + 4\delta \quad \text{and all } k \geq k_R. \tag{13.4.3}$$

This implies, by Exercise 13.3.16, that

$$(\gamma_k(-t)|\zeta')_{\gamma(0)} \geq R \quad \text{and} (\gamma_k(t)|\zeta''')_{\gamma(0)} \geq R \quad \text{for all } t \geq R + 4\delta \quad \text{and } k \geq k_R. \tag{13.4.4}$$

Consequently , for all $k \geq k_R$ and $t \geq R + 4\delta$, the points $\gamma_k(t)$ are contained in the Gromov neighborhood $V(\zeta'''; s)$, and the points $\gamma_k(-t)$ are contained in $V(\zeta'; s)$, where $s = s(R) = R - 4\delta - d(\beta(0), 1)$. (This follows because

the difference between the Gromov products $(\cdot|\cdot)_{\gamma(0)}$ and $(\cdot|\cdot)_1$ is bounded by $d(\beta(0), 1)$, by (13.3.3).)

By construction, $\xi \notin \{\zeta', \zeta'''\}$, so there exist $r, R \in \mathbb{N}$ so large that the Gromov neighborhood $V(\xi; r)$ intersects neither $V(\zeta'; s(R))$ nor $V(\zeta'''; s(R))$. Moreover, given $R \in \mathbb{N}$, we may choose r so large that $V(\xi; r)$ does not intersect the (word metric) ball of radius $R + 4\delta$ centered at $\beta(0)$. For any such r, the relations (13.4.3) and (13.4.4) ensure that *none* of the geodesic lines $\{\gamma_k\}_{k \geq k_R}$ *intersect* $V(\xi; r)$. This contradicts (13.4.2). $\square$

Proposition 13.4.2, together with the next two propositions, will imply that each group element $g \in \Gamma$ of infinite order has a well -defined "source" and "sink" on $\partial\Gamma$: in particular, the sequences $(g^n \cdot z)_{n \in \mathbb{N}}$ and $(g^{-n} \cdot z)_{n \in \mathbb{N}}$ converge to distinct points $\xi_+, \xi_- \in \partial\Gamma$, called the *poles* of g; furthermore, the homeomorphisms $\eta \mapsto g^n \cdot \eta$ of the boundary converge uniformly on $\partial\Gamma \setminus \{\xi_-\}$ to the constant mapping $\eta \mapsto \xi_+$ as $n \to \infty$.

Proposition 13.4.4 *For any element g of infinite order in Γ there is a g-invariant, quasi-geodesic line in the Cayley graph $G_{\Gamma;\mathbb{A}}$ on which g acts by translation. In particular, if α_g is a geodesic segment from 1 to g, then the bi-infinite path $\cup_{n \in \mathbb{Z}} g^n \cdot \alpha_g$ is quasi-geodesic.*

The proof will rely on the following variant of Exercise 13.1.4.

Exercise 13.4.5 Let $(\mathcal{X}, d)$ be a δ-hyperbolic metric space. Fix $R \gg \delta$, and let $x, x', y, y' \in \mathcal{X}$ be points such that $d(x, x') = 16R$ and $\max(d(x, y), d(x', y')) \leq R$. Fix geodesic segments β, γ with endpoints x, x' and y, y', respectively, and let b and c be the midpoints of β and γ. Show that there is a point $b' \in \beta$ such that

$$d(b, b') \leq 2R \quad \text{and} \quad d(b', c) \leq 2\delta.$$

HINT: Exercise 13.1.4 implies that $d(c, \beta) \leq 2\delta$. Now use the fact that β and γ are geodesics to show that a geodesic projection of c onto β cannot be farther than $2R$ from b.

Proof of Proposition 13.4.4. The path $\gamma := \cup_{n \in \mathbb{Z}} g^n \cdot \alpha_g$ is obviously invariant under left translation by g. Since left translation by any group element is an isometry of the Cayley graph, all of the segments $g^n \cdot \alpha_g$ have the same length, and so for any $n \in \mathbb{Z}$ and $m \in \mathbb{N}$,

$$\text{arc length} \cup_{i=n+1}^{n+m} g^i \cdot \alpha_g = m \times \text{arc length}(\alpha_g).$$

Thus, to show that γ is quasi-geodesic, it suffices to exhibit a constant $K \geq 1$ such that

$$K^{-1}m \leq |g^m| = d(1, g^m) \leq Km \quad \text{for every } m \in \mathbb{N}. \tag{13.4.5}$$

Because each $h \in \Gamma$ acts as an isometry on $G_{\Gamma;\mathbb{A}}$, the sequence $(|h^n|)_{n\in\mathbb{N}}$ is subadditive. Hence, by Lemma 3.2.1 the limit

$$\tau(h) := \lim_{n\to\infty} \frac{|h^n|}{n}, \tag{13.4.6}$$

called the *(asymptotic) translation length* of h, exists and is nonnegative. Therefore, proving (13.4.5) is equivalent to showing that any element g of infinite order in a hyperbolic group has *positive* translation length.

Fix an integer $R \gg \delta$ (eventually we will let $R \to \infty$), and let α be a geodesic segment in $G_{\Gamma;\mathbb{A}}$ of length $16R$, with endpoints $1, x \in \Gamma$ and midpoint $w \in \Gamma$. Each translate $g^n \cdot \alpha$ has length $16R$ and midpoint $g^n w \in \Gamma$. By Exercise 13.4.5, for any $n \geq 1$ such that $|g^n| \leq R$ there is a point y_n on α such that

$$d(y_n, w) \leq 2R \quad \text{and} \quad d(y_n, g^n \cdot w) \leq 2\delta,$$

and since each point of the Cayley graph is within distance 1 of a group element, it follows that for each n there exists $\tilde{y}_n \in \Gamma$ such that

$$d(\tilde{y}_n, w) \leq 2R \quad \text{and} \quad d(\tilde{y}_n, g^n \cdot w) \leq 2\delta + 1.$$

But the number of distinct vertices $h \in \Gamma$ within distance $2\delta + 1$ of the geodesic segment $\beta \subset \alpha$ with midpoint w and length $2R$ is bounded by $C(2R + 1)$, where C is the cardinality of the ball $\mathbb{B}_{2\delta+1}(1)$ in Γ. Since g has infinite order in Γ, the points $(g^n w)_{n\in\mathbb{N}}$ are distinct. Hence, there cannot be more than $C(2R+1)$ elements of the sequence $\{g^n\}_{n\in\mathbb{N}}$ within distance R of the group identity; in particular, there exists $1 \leq n \leq 2C(2R + 1)$ such that $|g^n| > R$. Since this holds for all R sufficiently large, it follows that

$$\tau(g) \geq \frac{1}{2C}.$$

□

Corollary 13.4.6 *For any element $g \in \Gamma$ of infinite order there are exactly two fixed points $\xi_+, \xi_- \in \partial\Gamma$ of the mapping $g : \partial\Gamma \to \partial\Gamma$ defined by $\eta \mapsto g \cdot \eta$. Every g-invariant quasi-geodesic has endpoints $\xi_\pm$, and consequently any two g-invariant quasi-geodesics are at finite Hausdorff distance. Furthermore, the two fixed points $\xi_\pm$ can be labeled so that for every $x \in \Gamma$ and every $\eta \in \partial\Gamma \setminus \{\xi_\pm\}$,*

$$\lim_{n\to\infty} g^n \cdot x = \lim_{n\to\infty} g^n \cdot \eta = \xi_+ \quad \textit{and} \tag{13.4.7}$$

$$\lim_{n\to-\infty} g^n \cdot x = \lim_{n\to-\infty} g^n \cdot \eta = \xi_-. \tag{13.4.8}$$

Proof. Proposition 13.4.4 implies that if α is a geodesic segment from 1 to g then the path $\beta := \cup_{n\in\mathbb{Z}} g^n \cdot \alpha$ is a g-invariant, quasi-geodesic line in the Cayley graph.

Clearly, the subsets

$$\beta_+ := \cup_{n\geq 0} g^n \cdot \alpha \quad \text{and}$$
$$\beta_- := \cup_{n\leq -1} g^n \cdot \alpha$$

are quasi-geodesic rays, so by Proposition 13.3.12 have well-defined limits $\xi_+, \xi_- \in \partial\Gamma$, and in particular, the sequences $(g^n)_{n\geq 0}$ and $(g^n)_{n\leq 0}$ converge to the points ξ_+ and ξ_-, respectively, in the Gromov topology. Because the mapping $x \mapsto g \cdot x$ is an isometry of the Cayley graph, for any $x \in \Gamma$,

$$d(g^n x, g^n) = d(x, 1) = |x| \quad \text{for all } n \in \mathbb{Z};$$

hence, by (13.3.1), $\lim_{n\to\pm\infty}(g^n x|g^n) = \infty$, and so the sequences $(g^n x)_{n\geq 0}$ and $(g^n x)_{n\leq 0}$ also converge to the points ξ_+ and ξ_-.

Now suppose that β' is another g-invariant quasi-geodesic line. For any point x on β' we have $g^n x \to \xi_\pm$ as $n \to \pm\infty$, so by g-invariance the points $\xi_\pm$ must be the endpoints of β'. Consequently, Proposition 13.3.15 implies that the geodesics β, β' must be at finite Hausdorff distance.

Finally, let $\eta \in \partial G \setminus \{\xi_+, \xi_-\}$. By Proposition 13.3.14, there is a geodesic ray γ that converges to η. By Exercise 13.3.16, the rays γ and β_+ separate at distance $\approx (\eta|\xi_+)_1$ from 1. Fix a point x near the separation point. Since $\lim_{n\to\infty} g^n \cdot x = \xi_+$, the separation points of the geodesic rays $g^n \cdot \gamma$ and $g^n \cdot \beta_+$ converge to ξ_+ as $n \to \infty$; therefore, we must have $\lim_{n\to\infty} g^n \cdot \eta = \xi_+$. The same argument shows that $\lim_{n\to-\infty} g^n \cdot \eta = \xi_-$. □

The *centralizer* of a group element $g \in \Gamma$ is the set $C(g)$ consisting of all $h \in \Gamma$ such that $gh = hg$. Clearly, $C(g)$ is a subgroup of Γ, and the cyclic group $\langle g \rangle := \{g^n\}_{n\in\mathbb{Z}}$ is a normal subgroup of $C(g)$.

Corollary 13.4.7 *If $g \in \Gamma$ is an element of infinite order then the centralizer of g is a finite extension of the infinite cyclic group $\mathbb{Z}$. Consequently, the quotient group $C(g)/\langle g \rangle$ is finite.*

This limits the types of abelian subgroups that can live inside hyperbolic groups: for instance, no hyperbolic group can contain a subgroup isomorphic to $\mathbb{Z}^2$.

Proof of Corollary 13.4.7. Suppose that $h \in C(g)$ and let $\beta = \cup_{n\in\mathbb{Z}} g^n \cdot \alpha$, where α is a geodesic segment in the Cayley graph with endpoints 1 and g^n. By Proposition 13.4.4, α is a g-invariant quasi-geodesic. Since left translation by h is an isometry, it follows that the image $h \cdot \beta$ is also a quasi-geodesic; and since h commutes with g, this quasi-geodesic is $g-$invariant. Hence, by Corollary 13.4.6, the Hausdorff distance D between β and $h \cdot \beta$ is finite. In particular, because every point on β is within distance $|g|$ of some g^n, the vertex $h \in h \cdot \beta$ must be within distance $D + |g|$ of some vertex $g^n \in \beta$, which implies that $hg^{-n} = g^{-n}h$ is an element of the ball $\mathbb{B}_{D+|g|}(1)$. It follows that the quotient group $C(g)/\langle g \rangle$ has at most $|\mathbb{B}_{D+|g|}(1)|$ elements. □

Corollary 13.4.8 *If* Γ *contains two elements* g, h *of infinite order such that the cyclic groups* $\{g^n\}_{n\in\mathbb{Z}}$ *and* $\{h^n\}_{n\in\mathbb{Z}}$ *have trivial intersection* $\{1\}$*, then for all sufficiently large integers* m, n *the subgroup of* Γ *generated by* g^m *and* h^n *is isomorphic to the free group* $\mathbb{F}_2$*. In this case, the the boundary* $\partial\mathbb{F}_2$ *maps injectively into the Gromov boundary* $\partial\Gamma$*, and so* $\partial\Gamma$ *is uncountable.*

Terminology: A hyperbolic group Γ that satisfies the hypothesis of Corollary 13.4.8 is called *nonelementary*. Since any such group contains a free subgroup, it must be nonamenable, and therefore subject to Kesten's Theorem (Theorem 5.1.5) and its consequences. In particular, a symmetric random walk on a nonelementary hyperbolic group Γ with finitely supported step distribution must have positive speed, positive Avez entropy, and nontrivial Poisson boundary. It is not difficult to show that a hyperbolic group is nonelementary if and only if its Gromov boundary has at least 3 points.

Proof of Corollary 13.4.8. Let $\xi_\pm$ be the poles of g and $\zeta_\pm$ those of h. We will first show that $\{\xi_+, \xi_-\} \cap \{\zeta_+, \zeta_-\} = \emptyset$. Suppose otherwise; then without loss of generality (by replacing one or both of the elements g, h with their inverses) we may assume that $\xi_+ = \zeta_+$. Let α_g, α_h be geodesic segments from 1 to g and 1 to h, respectively; then by Proposition 13.4.4, the quasi-geodesic lines $\beta_g = \cup_{n\in\mathbb{Z}} g^n \cdot \alpha_g$ and $\beta_h := \cup_{n\in\mathbb{Z}} h^n \cdot \alpha_h$ are g- and h-invariant, respectively. Furthermore,the quasi-geodesic rays $\beta_g^+ = \cup_{n\geq 0} g^n \cdot \alpha_g$ and $\beta_h^+ := \cup_{n\geq 0} h^n \cdot \alpha_h$ both converge to ξ_+. By Proposition 13.1.8, there exist *geodesic* rays γ_g, γ_h at finite Hausdorff distances from β_g^+, β_h^+, respectively. These geodesic rays both converge to ξ_+; hence, by Proposition 13.3.15, they are at finite Hausdorff distance, and so by the triangle inequality for the Hausdorff distance it follows that the quasi-geodesic rays β_g^+ and β_h^+ are at finite Hausdorff distance D. This implies, in particular, that the points $\{g^n\}_{n\geq 1}$ are all at distance $\leq D$ from β_h^+. Since each segment $h^n \cdot \alpha_h$ of β_h^+ has length $|h|$, it follows that for every $n \in \mathbb{N}$ there exists $m \geq 0$ such that

$$d(g^n, h^m) = d(1, g^{-n} h^m) \leq D + |h|,$$

and since both g, h have positive translation length, the ratios m/n of any two such integers is bounded above and below. But there are only finitely many elements of Γ within distance $D + |h|$ of the group identity. Therefore, there must be positive integers $m > m'$ and n, n' such that $g^{-n} h^m = g^{-n'} h^{m'}$, equivalently,

$$g^{n'-n} = h^{m'-m},$$

in contradiction to our hypothesis on the elements g, h.

Next, let A_+, A_-, B_+, B_- be open neighborhoods in $\partial\Gamma$ of the points $\xi_+, \xi_-, \zeta_+, \zeta_-$, respectively, whose closures $\bar{A}_+, \bar{A}_-, \bar{B}_+, \bar{B}_-$ are pairwise disjoint. By relation 6.4.9, if $m, n \in \mathbb{N}$ are sufficiently large, then

$$g^m(\bar{A}_+ \cup \bar{B}_- \cup \bar{B}_+) \subset A_+ \quad \text{and} \quad g^{-m}(\bar{A}_- \cup \bar{B}_- \cup \bar{B}_+) \subset A_-$$
$$h^n(\bar{B}_+ \cup \bar{A}_- \cup \bar{A}_+) \subset B_+ \quad \text{and} \quad h^{-n}(\bar{B}_- \cup \bar{A}_- \cup \bar{A}_+) \subset B_-. \tag{13.4.9}$$

Consequently, by Klein's Ping-Pong Lemma (Section 5.2), the subgroup of Γ generated by $a = g^m$ and $b = h^n$ is free.

To any finite reduced word w of length ≥ 1 in the letters $a^{\pm}, b^{\pm}$ associate a closed subset $F_w \subset \partial\Gamma$ as follows: if $w = av$ begins with the letter a (respectively, with a^{-1}, or b, or b^{-1}), set $F_w = w(A_+)$ (respectively, $F_w = w(A_-)$, or $F_w = w(B_+)$, or $F_w = w(B_-)$). Then the relations (13.4.9) imply, by induction on word length, that for any finite reduced words w, w',

(1) if w is a prefix of w' (i.e., $w' = wv$) then $F_{w'} \subset F_w$; and
(2) if neither w nor w' is a prefix of the other, then $F_w \cap F_{w'} = \emptyset$.

Since the sets F_w are nonempty and compact, the nesting property (1) ensures that for every infinite reduced word ω there is a point ξ_ω that is an element of F_w for every finite prefix w of ω, and property (2) guarantees that the mapping $\omega \mapsto \xi_\omega$ is one-to-one. □

Corollary 13.4.9 *If Γ is a nonelementary hyperbolic group, then the induced action of Γ on its Gromov boundary $\partial\Gamma$ is* transitive*, that is, for every point $\zeta \in \partial\Gamma$ the orbit $\Gamma \cdot \zeta$ is dense in $\partial\Gamma$.*

Proof. Fix points $\xi, \zeta \in \partial\Gamma$, and let $(g_n)_{n\in\mathbb{N}}$ be a sequence of group elements that converge to ξ in the Gromov topology. By Proposition 13.4.2, there exist a subsequence $(g_m)_{m\in\mathbb{N}}$ and a point $\eta \neq \xi$ such that $\lim_{m\to\infty} g_m \cdot \zeta = \xi$ for every $\zeta \neq \eta$. Thus, the orbit $\Gamma \cdot \zeta$ accumulates at ξ except possibly when $\zeta = \eta$.

Since Γ is nonelementary, there exist elements $g, h \in \Gamma$ of infinite order whose poles $\alpha_\pm, \beta_\pm$ are distinct. Hence, for at least one of these (say g), the poles α_+, α_- are contained in $\partial\Gamma \setminus \{\eta\}$. But this implies that $g^n \cdot \eta \to \alpha_+ \neq \eta$, and so for some $m \in \mathbb{N}$ it must be that $g^m \cdot \eta \neq \eta$, and so $\lim_{n\to\infty} g_n g^m \cdot \eta = \xi$. □

13.5 Random Walks on Hyperbolic Groups

Assumption 13.5.1 *Assume in this section that $(X_n)_{n\geq 0}$ is an irreducible random walk on a nonelementary hyperbolic group Γ with step distribution μ.*

Nonelementary hyperbolic groups are nonamenable, so the random walk $(X_n)_{n\geq 0}$ is transient, by virtue of irreducibility. By Corollary 13.3.18 and Proposition 13.3.10, the Gromov boundary $\partial\Gamma$ is compact and metrizable in the Gromov topology, and the group Γ acts by homeomorphisms on $\partial\Gamma$, so by Proposition 11.2.4, there is a μ-stationary Borel probability measure λ on $\partial\Gamma$. Our primary objective in this section is to establish the following theorem.

Theorem 13.5.2 *The stationary measure* λ *is unique. Furthermore, if the random walk has finite Avez entropy and if its step distribution* μ *has finite first moment* $\sum_{g\in\Gamma}|g|\mu(g)$ *then the pair* $(\partial\Gamma, \lambda)$ *is a Furstenberg-Poisson boundary.*

The substantive part of this theorem — that $(\partial\Gamma, \lambda)$ is a *Poisson* boundary — is due to Kaimanovich [70]. By Corollary 13.4.9, the action of Γ on its Gromov boundary $\partial\Gamma$ is transitive, so by Proposition 11.3.4, there is only one μ-stationary probability measure λ on $\partial\Gamma$. This measure λ is nonatomic, and it assigns positive probability to every nonempty open neighborhood of $\partial\Gamma$.

To prove that the pair $(\partial\Gamma, \lambda)$ is a Furstenberg-Poisson boundary, we must first show that it is a μ-boundary in the sense of Definition 11.4.6.

Proposition 13.5.3 *With* P^1*-probability one, the sequence* $(X_n)_{n\in\mathbb{N}}$ *converges in the Gromov topology to a (random) point* $X_\infty \in \partial\Gamma$ *whose distribution is* λ*. Consequently, the pair* $(\partial\Gamma, \lambda)$ *is a* μ*-boundary.*

Proof. By Lemma 11.4.8, with probability one the measures $X_n \cdot \lambda$ converge weakly. We must show that with probability one the limit is a point mass at a random point X_∞.

By hypothesis the group Γ is nonelementary, and hence nonamenable, so the random walk $(X_n)_{n\in\mathbb{N}}$, being irreducible, is transient. Hence, since $\partial\Gamma$ is compact in the Gromov topology, with probability one every subsequence of $(X_n)_{n\in\mathbb{N}}$ has a subsequence $(X_k)_{k\in\mathbb{N}}$ that converges to a point Z in $\partial\Gamma$. By Proposition 13.4.2, $(X_k)_{k\in\mathbb{N}}$ has a subsequence $(X_j)_{j\in\mathbb{N}}$ such that the homeomorphisms $\zeta \mapsto X_j \cdot \zeta$ of the boundary converge to the constant mapping $\zeta \mapsto Z$ on all but at most one point of $\partial\Gamma$. Therefore, since λ is nonatomic, the measures $X_j \cdot \lambda$ must converge weakly to the point mass at Z. Since the full sequence $(X_n \cdot \lambda)_{n\in\mathbb{N}}$ converges weakly, it follows that there can be only one subsequential limit point $Z := X_\infty$ for the sequence $(X_n)_{n\in\mathbb{N}}$, and that $X_n \cdot \lambda \to \delta_{X_\infty}$ weakly. This proves that the pair $(\partial\Gamma, \lambda)$ is a μ-boundary, and, by Lemma 11.4.4, that λ is the distribution of the limit point X_∞. $\square$

Remark 13.5.4 If $(X_n)_{n\geq 0}$ is a random walk with step distribution μ and initial point 1 then for any $x \in \Gamma$ the translated sequence $(X'_n)_{n\geq 0} := (xX_n)_{n\geq 0}$ is a random walk with step distribution μ and initial point $X'_0 = x$. Since left translation by x induces a continuous mapping of the Gromov boundary, it follows from Proposition 13.5.3 that with P^x-probability one the sequence $(X_n)_{n\geq 0}$ converges to a random point $X_\infty \in \partial\Gamma$ with distribution $x \cdot \lambda$.

The almost sure existence of the limit X_∞ is the key to establishing the localization property of Corollary 12.2.5, as this forces the random walk path to travel roughly along the geodesic ray with endpoint X_∞ at a positive speed $\ell = a.s. - \lim_{n\to\infty} |X_n|/n$. The next proposition makes this precise.

Theorem 13.5.5 (Geodesic Tracking) *Assume that the step distribution* μ *of the random walk* $(X_n)_{n\geq 0}$ *has finite first moment. Denote by* $X_\infty := \lim_{n\to\infty} X_n$ *the almost sure limit of the sequence* $(X_n)_{n\geq 0}$*, and by* $\ell \in (0, \infty)$ *the speed of the*

random walk. Then for any geodesic ray γ that converges to X_∞,

$$\lim_{n\to\infty} \frac{1}{n} d(X_n, \gamma(\ell n)) = 0 \quad P^x - \textit{almost surely.} \tag{13.5.1}$$

Remark 13.5.6 The Subadditive Ergodic Theorem (cf. Corollary 3.3.2) implies that an irreducible random walk whose step distribution has finite first moment must have finite speed

$$\lim_{n\to\infty} \frac{|X_n|}{n} = \ell \quad \text{almost surely.} \tag{13.5.2}$$

The speed ℓ must be *positive*, because otherwise the return probabilities $P\{X_{2n} = 1\}$ would not decay exponentially, in contradiction to Day's Theorem 5.1.6.

The proof of Proposition 13.5.5 will hinge on the following lemma, which in part quantifies the convergence of the random walk to its boundary limit X_∞.

Lemma 13.5.7 *Under the hypotheses of Theorem 13.5.5, for every $\varepsilon > 0$, with P-probability one, the Gromov products $(X_n|X_\infty)$ satisfy*

$$(X_n|X_\infty)_1 \geq (\ell - \varepsilon)n \quad \textit{eventually.} \tag{13.5.3}$$

Proof. Since the step distribution has finite first moment, the Borel-Cantelli Lemma (see Corollary A.5.1) implies that for any $\varepsilon > 0$ and any $k = 1, 2, 3, \cdots$,

$$\lim_{n\to\infty} d(X_n, X_{n+k})/n = 0. \tag{13.5.4}$$

Together with relation (13.5.2), this implies that for any $\varepsilon > 0$ and any $k \in \mathbb{N}$,

$$(X_n|X_{n+k})_1 \geq (\ell - \varepsilon)n \tag{13.5.5}$$

eventually, with probability one. Fix $k \in \mathbb{N}$ large enough that $(\ell - \varepsilon)k > 4\delta$, and suppose that for some sample path $(X_n)_{n\geq 0}$ of the random walk the relation (13.5.5) holds for all $n \geq n_*$. Then by the ultrametric property (13.3.6) and induction on j, for all $n \geq n_*$ and all $j = 1, 2, \cdots$ we have

$$\begin{aligned}
(X_n|X_{n+jk})_1 &\geq \min((X_n|X_{n+k})_1, (X_{n+k}|X_{n+jk})_1) - 4\delta \\
&\geq \min((\ell - \varepsilon)n, (\ell - \varepsilon)(n + k) - 4\delta) - 4\delta \\
&\geq (\ell - \varepsilon)n - 4\delta.
\end{aligned}$$

Since $\lim_{k\to\infty} X_{n+k} = X_\infty$ almost surely, the definition (13.3.8) implies (13.5.3), with ε replaced by 2ε. □

Proof of Theorem 13.5.5. For each n let α_n be a geodesic segment from 1 to X_n, and as usual let α_n be parametrized by arc length. For all large n the length of α_n is at least $\ell n - \varepsilon n$, by (13.5.2), so there is a point $Y_n = \alpha_n(\ell n - 2\varepsilon n)$ on α_n at distance $\ell n - 2\varepsilon n$ from the initial point 1. Relation (13.5.2) also implies that with probability one, for all large n the distance between Y_n and the terminal endpoint X_n satisfies

$$\varepsilon n \leq d(Y_n, X_n) \leq 3\varepsilon n. \tag{13.5.6}$$

By Lemma 13.5.7 and Exercise 13.3.16,

$$d(Y_n, \gamma(\ell n - 2\varepsilon n)) \leq 2\delta + 1, \tag{13.5.7}$$

for all large n, with probability one, and therefore, by the triangle inequality,

$$d(X_n, \gamma(\ell n)) \leq 5\varepsilon n + 2\delta + 1 \quad \text{almost surely.} \tag{13.5.8}$$

Finally, since $\varepsilon > 0$ is arbitrary, the relation (13.5.1) follows. □

Proof of Theorem 13.5.2. To complete the proof of Theorem 13.5.2, we must show that the measure space $(\partial\Gamma, \lambda)$ is a Poisson boundary. For this we will apply the localization criterion of Corollary 12.2.5. This criterion requires measurable mappings $B_n : \partial\Gamma \to \mathfrak{S}$, where $\mathfrak{S}$ is the space of finite subsets of Γ, such that conditions (12.2.11) and (12.2.12) are satisfied. Here we will use (rough) balls $B_{n,\varepsilon}(\xi)$ of radii $n\varepsilon$ centered at points at distance ℓn along geodesic rays converging to boundary points $\xi \in \partial\Gamma$. Given a boundary point ξ, there will, in general, be many geodesic rays with initial point 1 that converge to ξ; however, by Exercise 13.3.16, for any two such rays γ_1, γ_2,

$$d(\gamma_1(t), \gamma_2(t)) \leq 2\delta + 1 \quad \text{for all } t \geq 0. \tag{13.5.9}$$

Fix $\varepsilon > 0$, and define $B_{n,\varepsilon}(\xi)$ to be the set of all $g \in \Gamma$ such that for *some* geodesic ray γ with initial point 1 that converges to ξ,

$$d(g, \gamma(\ell n)) \leq \varepsilon n. \tag{13.5.10}$$

Then by (13.5.9),

$$|B_{n,\varepsilon}(\xi)| \leq |\mathbb{A}|^{n\varepsilon + 2\delta + 1}, \tag{13.5.11}$$

so the requirement (12.2.11) holds (with a possibly different choice of $\varepsilon > 0$). The requirement (12.2.12) is a direct consequence of Theorem 13.5.5. □

13.6 Cannon's Lemma

Any two geodesic segments in a δ-hyperbolic space with the same endpoints must be at Hausdorff distance $\leq \delta$, by the thin triangle property — see the remarks preceding Lemma 13.2.6. In a Cayley graph of a hyperbolic group, there are only finitely many paths through any ball of radius R, so there are only finitely many possible "local routes" for geodesics heading toward the same terminal point. This has far-reaching consequences. The key to unlocking these is the notion of a *cone type*, first introduced (in this context) by Cannon [21].

Definition 13.6.1 Let Γ be a finitely generated group with finite, symmetric generating set $\mathbb{A}$ and Cayley graph $G_{\Gamma;\mathbb{A}}$. For any group element g, the *cone* $C(g)$ based at g is the subgraph of $G_{\Gamma;\mathbb{A}}$ whose vertex set is the collection of all $h \in \Gamma$ such that some geodesic segment in $G_{\Gamma;\mathbb{A}}$ with endpoints 1 and h passes through g. The *cone type* $\tau(g)$ is the set of all $h \in \Gamma$ such that $gh \in C(g)$. The *neighborhood type* $\nu_m(g)$ of g at depth $m \in \mathbb{N}$ is the set of all group elements h in the ball $\mathbb{B}_m(1)$ such that $|gh| < |g|$.

Exercise 13.6.2 (A) Show that the Cayley graph of the free group $\mathbb{F}_2$ with the standard generating set $\mathbb{A} = \{a, b, a^{-1}, b^{-1}\}$ has five distinct cone types. (B) Show that Cayley graph of the group $\mathbb{F}_2 \times \mathbb{Z}_2$ with the standard generating set $\mathbb{A} = \{a, b, a^{-1}, b^{-1}\} \times \{0, 1\}$ has ten distinct cone types. (Note: This group is hyperbolic, as it has $\mathbb{F}_2$ as a subgroup of index 2 — see the remarks following the statement of Proposition 13.1.9.)

Lemma 13.6.3 (Cannon [21]) *If the Cayley graph $G_{\Gamma;\mathbb{A}}$ is δ-hyperbolic then the cone type $\tau(g)$ of any group element g is determined by the neighborhood type $\nu_m(g)$, where m is any integer larger than $2\delta + 1$. Consequently, the group Γ has only finitely many cone types.*

Proof. The cone $C(1)$ based at the group identity is the entire Cayley graph, and the neighborhood type $\nu_m(1)$ is the null set. No other group element has the same cone type or neighborhood type. Therefore, in proving the lemma we can ignore the group element $g = 1$.

Fix $m > 2\delta + 1$, and let $g_1, g_2 \in \Gamma$ be two elements such that $\nu_m(g_1) = \nu_m(g_2)$. To establish that g_1 and g_2 have the same cone type, we will prove by induction on n that for any group element h of word length $|h| = n$, either $h \in C(g_1) \cap C(g_2)$ or $h \in C(g_2)^c \cap C(g_2)^c$. The case $n = 0$ is trivial, because the only element $h \in \Gamma$ such that $|h| = 0$ is the identity $h = 1$, and clearly $1 \in C(g_1) \cap C(g_2)$. Suppose, then, that the assertion is true for all h' such that $|h'| \leq n$. We will prove that if $h \in C(g_1)$ has word length $|h| = n + 1$, then $h \in C(g_2)$.

If $h \in C(g_1)$, then there is a geodesic segment from 1 to $g_1 h$ that passes through g_1. This geodesic segment is the concatenation $\alpha_1 * (g_1 \cdot \beta)$ of a geodesic segment α_1 from 1 to g_1 with the translate $g_1 \cdot \beta$ of a geodesic segment β from 1 to h. We will show that if α_2 is any geodesic segment from 1 to g_2 then the concatenation

$\alpha_2 * (g_2 \cdot \beta)$ of α_2 with the translate $g_2 \cdot \beta$ is a geodesic segment. Since the terminal endpoint of $\alpha_2 * (g_2 \cdot \beta)$ is $g_2 h$, this will prove that $h \in C(g_2)$.

Let h' be the vertex on β adjacent to h (thus, $|h'| = |h| - 1 = n$), and let β' be the sub-segment of β from 1 to h'. Every vertex h'' on β' has word length $|h''| \leq n$ and is an element of $C(g_1)$, since β' is geodesic. Consequently, by the induction hypothesis, every vertex h'' on β' is also an element of $C(g_2)$, and so the concatenation $\alpha_2 * (g_2 \cdot \beta')$ is geodesic, with terminal endpoint $g_2 h'$.

Suppose now that $\alpha_2 * (g_2 \cdot \beta)$ is *not* geodesic; we will show that this leads to a contradiction. If $\alpha_2 * (g_2 \cdot \beta)$ is not geodesic then there must be a geodesic segment γ from 1 to $g_2 h$ shorter than $\alpha_2 * (g_2 \cdot \beta)$. Hence, the path $\gamma' = \gamma * e(g_2 h, g_2 h')$ obtained by appending the edge $e(g_2 h, g_2 h')$ to γ has length no greater than that of $\alpha_2 * (g_2 \cdot \beta')$. Since the segment $\alpha_2 * (g_2 \cdot \beta')$ *is* geodesic, it follows that

$$\text{length}(\gamma') = \text{length}(\alpha_2 * (g_2 \cdot \beta'))$$

and so both paths are geodesic segments from 1 to $g_2 h'$. Let x be the vertex on γ' at distance $n + 1$ from $g_2 h'$ (and therefore at distance n from $g_2 h$).

Claim 13.6.4 $d(x, g_2) \leq 2\delta + 1$.

Proof of the Claim. Because any two geodesic segments with the same endpoints are within Hausdorff distance δ, there must be a point y on $\alpha_2 * (g_2 \cdot \beta')$ such that $d(x, y) \leq \delta$. This point y must be within distance $\delta + 1$ of g_2, because otherwise one of the paths γ' or $\alpha_2 * (g_2 \cdot \beta')$ from 1 to $g_2 h'$ could be shortened (by "crossing over" from one path to the other). □

By hypothesis, the path γ is shorter than $\alpha_2 * (g_2 \cdot \beta)$. Since the vertices x on γ and g_2 on $\alpha_2 * (g_2 \cdot \beta)$ are both at distance n from the common endpoint $g_2 h$, it follows that $|x| < |g_2|$, or equivalently, $|g_2 z| < |g_2|$, where $z = g_2^{-1} x$. By the Claim, $|z| \leq 2\delta + 1$, and so by definition $z \in \nu_m(g_2)$ for any $m \geq 2\delta + 1$. But for some $m \geq 2\delta + 1$ the neighborhood types $\nu_m(g_2)$ and $\nu_m(g_1)$ are the same; hence, $z \in \nu_m(g_1)$, and so in particular $|g_1 z| < |g_1|$.

Let γ'' be the sub-segment of γ running from $x = g_2 z$ to $g_2 h$, and let $\vartheta = (z^{-1} g_2^{-1}) \cdot \gamma''$ be its translation by x^{-1}, running from 1 to $z^{-1} g_2^{-1} g_2 h = z^{-1} h$. By construction, the path ϑ has length n. Therefore, if ψ is any geodesic segment from 1 to $g_1 z$ then the concatenation $\psi * ((g_1 z) \cdot \vartheta)$ has length $|g_1 z| + n < |g_1| + n = |g_1 h|$. This is impossible, because the terminal endpoint of $\psi * ((g_1 z) \cdot \vartheta)$ is $g_1 h$. □

A non-backtracking path $\alpha = (g_0, g_1, \cdots, g_n)$ in a Cayley graph $G_{\Gamma;\mathbb{A}}$ with initial vertex $g_0 = 1$ determines — and is determined by — the finite reduced word $(a_1, a_2, \cdots, a_n)$ in the generators defined by the requirement that $g_m = g_{m-1} a_m$ for every $1 \leq m \leq n$. Not every non-backtracking path is geodesic, so not every reduced word corresponds to a geodesic segment; those that do will be called *geodesic words*. Cannon's Lemma implies that the set of all finite geodesic words is a *regular language*, that is, it is the set of all finite words generated be a *finite-state automaton*.

Definition 13.6.5 A *labeled digraph* on an alphabet A is a pair $(V, \mathcal{E}))$ consisting of a *vertex set* V and a set $\mathcal{E} \subset V \times V \times A$ of *labeled, directed edges*. For any edge $\mathbf{e} = (v, w, a)$, the *initial* and *terminal* endpoints are v and w, respectively, and the *label* is the letter a. A *path* in a directed graph $G = (V, \mathcal{E}))$ is a sequence of directed edges in which the initial vertex of each edge is the terminal vertex of the preceding edge.

Observe that for any given pair v, w of edges, there can be several directed edges with initial and terminal vertices v and w, but at most one with a given label $a \in \mathbb{A}$.

Definition 13.6.6 A *finite-state automaton* on an alphabet A is a finite, labeled digraph $G = (V, \mathcal{E})$ with label set A, a distinguished vertex s (the *start vertex*), and a subset $F \subset V$ (the *accept states*). Each finite path β in G determines a finite word $w(\beta)$ with letters in the alphabet A: this word $w(\beta)$ is obtained by reading off the labels of successive edges crossed by the path β. The *regular language* generated by the automaton (G, s, F) is the set of all words $w(\beta)$ produced by paths β with initial vertex s and terminal vertex in the set F.

There are a number of variations of this definition: see Hopcroft and Ullman [66] for a detailed discussion.

Corollary 13.6.7 *If Γ is a hyperbolic group with finite, symmetric generating set $\mathbb{A}$ then the set of all finite geodesic words for the Cayley graph $G_{\Gamma;\mathbb{A}}$ is a regular language.*

Proof. A finite-state automaton $\mathcal{A} = \mathcal{A}(G_{\Gamma;\mathbb{A}})$ that generates the language of geodesic words can be constructed as follows. Let V, the vertex set, be the set of cone types. For any pair of cone types τ, τ' and any generator $a \in \mathbb{A}$, the triple (τ, τ', a) is a directed edge if and only if there exists $g \in \Gamma$ with cone type $\tau(g) = \tau$ such that $a \in \tau$ and $\tau' = \tau(ga)$. This specification does not depend on the choice of representative g for the cone type, by Definition 13.6.1. Finally, let the start state s be the cone type $\tau(1)$ of the group identity. Clearly, any finite word accepted by the automaton represents a geodesic segment in the Cayley graph $G_{\Gamma;\mathbb{A}}$ with initial endpoint 1, and conversely, any geodesic word represented by the automaton is accepted by the automaton. Cannon's Lemma ensures that the vertex set V is finite, so $\mathcal{A} = \mathcal{A}(G_{\Gamma;\mathbb{A}})$ is in fact a finite-state automaton. $\square$

Definition 13.6.8 The automaton $\mathcal{A} = \mathcal{A}(G_{\Gamma;\mathbb{A}})$ described above is the *Cannon automaton* associated with the Cayley graph $G_{\Gamma;\mathbb{A}}$, and the labeled digraph $\mathbf{G} = (V, \mathcal{E})$ on which the automaton runs is the *cone type digraph.* A cone type $\tau \in V$ is *recurrent* if there is a finite path in $\mathbf{G}$ of length at least 1 that begins and ends at τ. A cone type that is not recurrent is *transient*.

Exercise 13.6.9 (A) Show that if $R < \infty$ is sufficiently large then every ball of radius R in the Cayley graph contains a vertex whose cone type is recurrent. (B) Show that for any recurrent cone type τ and any group element g with cone type $\tau(g) = \tau$ there is a geodesic ray in the Cayley graph with initial vertex g that lies entirely in the cone $C(g)$.

HINT for Part (A): It suffices to show this for balls whose centers are at distance $\geq R$ from the group identity 1. The geodesic segment from 1 to the center x of such a ball $\mathbb{B}_R(x)$ must visit at least R successive vertices in $\mathbb{B}_R(x)$. Show that if R is sufficiently large then not all of the vertices on this geodesic segment could have transient cone types.

The Cannon automaton, in addition to determining the language of finite geodesic words, also determines the set of (infinite) geodesic rays in the Cayley graph with initial vertex 1, since these are precisely the paths for which every initial finite segment is geodesic. In particular,

(a) for any infinite path π in the digraph $\mathbf{G}$ with initial vertex $\tau(1)$ (the cone type of the group identity), the corresponding infinite path γ in the Cayley graph $G_{\Gamma;\mathbb{A}}$ with initial vertex 1 that follows the sequence of edges in $G_{\Gamma;\mathbb{A}}$ determined by the labels of the corresponding edges on π is a geodesic ray; and
(b) for any geodesic ray γ in $G_{\Gamma;\mathbb{A}}$ with initial vertex 1, there is a corresponding infinite path π in $\mathbf{G}$ with initial vertex $\tau(1)$ that follows the sequence of edges determined by the labels $a_i \in \mathbb{A}$ of the successive edges in $G_{\Gamma;\mathbb{A}}$ crossed by γ. This path π can only make finitely many visits to transient vertices. (Exercise: Why?)

Remark 13.6.10 Cannon used Lemma 13.6.3 to prove a more subtle result than Corollary 13.6.7. Fix any total order $\prec$ on the generating set $\mathbb{A}$, and for each group element $g \in \Gamma$ let $w(g)$ be the *lexicographically smallest* geodesic word, relative to the order $\prec$, representing a geodesic segment from 1 to g. Cannon's theorem asserts that for any choice of the total order $\prec$, the set $\{w(g) \, : \, g \in \Gamma\}$ is a regular language. See [21], or alternatively [19], Theorem 3.2.2 for the proof.

13.7 Random Walks: Cone Points

According to Theorem 13.5.5, an irreducible random walk $(X_n)_{n\geq 0}$ on a nonelementary hyperbolic group whose step distribution has finite first moment must almost surely "track" a (random) geodesic ray γ in the sense that $d(X_n, \gamma) = o(n)$. In this section, we will use Cannon's Lemma to show that the sequence $(X_n)_{n\geq 0}$ must actually visit any geodesic ray γ that converges to the boundary point X_∞ infinitely often.

Proposition 13.7.1 *Let $(X_n)_{n\geq 0}$ be an irreducible random walk on a nonelementary hyperbolic group whose step distribution has finite first moment, and let $X_\infty = \lim_{n\to\infty} X_n$ be its P-almost sure Gromov limit. Then for any geodesic ray γ with boundary limit X_∞,*

$$\sum_{n=1}^{\infty} \mathbf{1}\{X_n \in \gamma\} = \infty \quad \textit{almost surely } (P). \tag{13.7.1}$$

Exercise 13.7.2 Show that it suffices to prove Proposition 13.7.1 for random walks whose step distributions assign positive probability to every element of the generating set $\mathbb{A}$.

HINT: Replace the step distribution μ by $n^{-1}\sum_{i=1}^{n}\left(\frac{1}{2}(\mu+\delta_1)\right)^{*i}$ for some $n \in \mathbb{N}$.

Assumption 13.7.3 *Assume for the remainder of this section that the step distribution μ gives positive probability to every $a \in \mathbb{A}$. Assume also that the Cayley graph $G_{\Gamma;\mathbb{A}}$ is δ-hyperbolic, and let $\mathcal{A} = \mathcal{A}(G_{\Gamma;\mathbb{A}})$ be the* Cannon automaton *of the Cayley graph.*

If $x \in \Gamma$ has recurrent cone type $\tau(x)$, then by Cannon's Lemma (cf. Exercise 13.6.9 (B)) there is an infinite geodesic ray γ with initial vertex 1 that passes through x and infinitely many other vertices with cone type $\tau(x)$. Let $\xi \in \partial\Gamma$ be the Gromov limit of this geodesic ray. If $r > |x| + 4\delta$ then any geodesic ray γ' with initial vertex 1 that converges to a point in the Gromov neighborhood $V(\xi; r)$ must pass through the ball $\mathbb{B}_{2\delta+1}(x)$ of radius $2\delta + 1$ centered at x (cf. Exercise 13.3.16), although it need not pass through the vertex x itself. For this reason, we define the *fat cone*

$$C^+(x) := \bigcup_{x' \in \mathbb{B}_{2\delta+1}(x)} C(x') \tag{13.7.2}$$

to be the union of all cones based at points x' in the ball $\mathbb{B}_{2\delta+1}(x)$. Define the *fat cone type* $\tau^+(x)$ of x to be the set of all pairs $(h, h') \in \mathbb{B}_{2\delta+1} \times \Gamma$ such that $xhh' \in C(xh)$. By Cannon's Lemma, the cone type of any element is determined by its m-neighborhood type, for any $m \geq 2\delta + 1$; consequently, the fat cone type is determined by the $2m$-neighborhood type, for any $m \geq 2\delta + 1$, and so there are only finitely many fat cone types.

Call a fat cone type τ^+ *recurrent* if for some (and hence every) $x \in \Gamma$ with fat cone type $\tau^+(x) = \tau^+$, the (ordinary) cone type $\tau(x)$ is recurrent. If $\tau^+(x)$ is recurrent, then there is a geodesic ray starting at 1 that passes through x and then remains in the cone $C(x) \subset C^+(x)$. The limit $\xi \in \partial\Gamma$ of this geodesic is contained in the set $\partial C^+(x) \subset \partial\Gamma$ consisting of all limits of geodesic rays that lie entirely in $C^+(x)$. Since any geodesic that converges to a point ξ' in the Gromov neighborhood $V(\xi; |x|+4\delta)$ must pass through the ball $\mathbb{B}_{2\delta+1}(x)$, the set $C^+(x) \cup \partial C^+(x)$ contains a nonempty Gromov-open set.

Proposition 13.7.4 (Haissinsky, Mathieu, Müller[63]) *For each recurrent fat cone type τ^+ there exists a scalar $p(\tau^+) > 0$ such that for any group element $x \in \Gamma$ with fat cone type $\tau^+(x) = \tau^+$,*

$$P^x\left\{X_n \in C^+(x) \text{ for every } n \geq 0\right\} = p(\tau^+). \tag{13.7.3}$$

Proof. If $x, x' \in \Gamma$ are two group elements with the same fat cone type, then for any sequence $g_1, g_2, \cdots$ in Γ,

$$xg_1g_2\cdots g_n \in C^+(x) \;\; \forall\, n \geq 0 \quad \Longleftrightarrow \quad x'g_1g_2\cdots g_n \in C^+(x') \;\; \forall\, n \geq 0. \tag{13.7.4}$$

Consequently,

$$P^x\left\{X_n \in C^+(x) \text{ for every } n \geq 0\right\} = P^{x'}\left\{X_n \in C^+(x') \text{ for every } n \geq 0\right\}.$$

By Proposition 13.5.3, with P-probability one the sequence $(X_n)_{n\in\mathbb{N}}$ converges in the Gromov topology to a (random) point $X_\infty \in \partial\Gamma$ whose distribution λ of the limit point X_∞ is nonatomic and assigns positive probability to every nonempty open subset of $\partial\Gamma$. For any $x \in \Gamma$ whose fat cone type is recurrent, the Gromov boundary $\partial C^+(x)$ contains a nonempty open subset of $\partial\Gamma$; therefore,

$$P\left\{X_\infty \in \partial C^+(x)\right\} = P\left\{X_n \in C^+(x) \;\text{ for all large } n \in \mathbb{N}\right\} > 0.$$

Consequently, for some $m \in \mathbb{N}$ there is positive P-probability that $X_n \in C^+(x)$ for all $n \geq m$, and hence, since Γ is countable, there is some $y \in C^+(x)$ such that

$$P\left\{X_m = y \;\text{ and }\; X_n \in C^+(x) \;\text{ for all } n \geq m\right\} > 0.$$

But by the Markov property,

$$\frac{P\left\{X_m = y \;\text{ and }\; X_n \in C^+(x) \;\text{ for all } n \geq m\right\}}{P\left\{X_m = y\right\}} = P^y\left\{X_n \in C^+(x) \text{ for all } n \geq 0\right\},$$

so there is positive probability that a random walk with step distribution μ starting at $X_0 = y$ will remain in $C^+(x)$ forever.

Since $y \in C^+(x)$, there must be some $z \in \mathbb{B}_{2\delta+1}(x)$ such that some geodesic segment β from z to y lies entirely in $C^+(x)$. Because the step distribution μ of the random walk gives positive probability to every generator $a \in \mathbb{A}$, there is positive probability that a random walk started at $X_0 = x$ will first follow a geodesic path in the ball $\mathbb{B}_{2\delta+1}$ from x to z, then follow the geodesic segment β from z to y. Thus, by the Markov property,

$$P^x\left\{X_n \in C^+(x) \;\text{ for all } n \geq 0\right\} > 0.$$

□

By Exercise 13.6.9(A) there exists $R < \infty$ such that every ball of radius R in the Cayley graph contains a vertex whose cone type is recurrent. Thus, regardless of its starting point, a random walk whose step distribution charges every generator will visit such a vertex x after having made a complete tour of the surrounding ball $\mathbb{B}_{2\delta+1}(x)$ within a short time. For any $k \in \mathbb{N}$ let m_k be the length of the shortest

nearest-neighbor path starting and ending at 1 that visits every $y \in \mathbb{B}_k(1)$, and define

$$\vartheta_0 := \min \Big\{ n \geq 1 + m_{2\delta+1} \; : \; \tau^+(X_n) \;\text{ is recurrent and }\; \{X_{n-j}\}_{j \leq m_{2\delta+1}} = \mathbb{B}_{2\delta+1}(X_n) \Big\} . \tag{13.7.5}$$

Lemma 13.7.5 *If the step distribution μ gives positive probability to every generator then there exist $r < 1$ and $C < \infty$ such that for any initial state $x \in \Gamma$ and all $n \in \mathbb{N}$,*

$$P^x \{\vartheta_0 \geq n\} \leq Cr^n. \tag{13.7.6}$$

Proof. Let $p = P^x(G_x)$, where G_x is the event that the random walk makes two consecutive tours of the ball $\mathbb{B}_{R+2\delta+1}(x)$ in its first $m := 2R + 2\delta + 1$ steps. This probability does not depend on x, and is strictly positive, since by hypothesis each nearest-neighbor step has positive probability. Because every ball of radius R contains a group element with recurrent cone type, the event G_x is contained in the event that a random walk started at x will make its first visit to a group element y with recurrent cone type and then make a complete tour of the ball $B_{2\delta+1}(y)$. Hence, by the Markov property and a routine induction,

$$P^x \{\vartheta_0 \geq mn\} \leq (1-p)^n \quad \text{for every } n \in \mathbb{N}.$$

This implies (13.7.6), with $r = (1-p)^{1/m}$ and $C = (1-p)^{-2/m}$. □

The random variable ϑ_0 is a stopping time for the random walk, because for any n the event that $\vartheta_0 = n$ is completely determined by the first n steps, and hence is an element of the σ-algebra $\sigma(X_i)_{i \leq n}$. Similarly, each of the random times ϑ_m defined inductively by

$$\vartheta_{m+1} = \min\{n > \vartheta_m + 1 + m_{2\delta+1} \; : \; \tau^+(X_n) \;\text{ is recurrent and }\; \{X_{n-j}\}_{j \leq m_{2\delta+1}} = \mathbb{B}_{2\delta+1}(X_n)\} \tag{13.7.7}$$

is a stopping time; these random variables are almost surely finite, because by Lemma 13.7.5 and the Markov property, for any $m, n, k \in \mathbb{Z}_+$,

$$P^x \{\vartheta_m = n \;\text{ and }\; \vartheta_{m+1} \geq n + k\} \leq P^x \{\vartheta_m = n\} (Cr^k). \tag{13.7.8}$$

At each time ϑ_m, there is, by Proposition 13.7.4, a positive (conditional) probability that the random walk will remain in the fat cone $C^+(X_{\vartheta_m})$ forever after. The next order of business is to show that for infinitely many $m \in \mathbb{N}$ this will occur. To this end, define random variables $\eta_0 < \eta_1 < \cdots$ inductively by $\eta_0 = \vartheta_0$ and

$$\eta_{k+1} = \min\left\{\vartheta_m > \eta_k \;:\; X_{\vartheta_m+n} \in C^+(X_{\vartheta_m}) \text{ for all } n \geq 0\right\}. \tag{13.7.9}$$

Unlike the random variables ϑ_m, these are not stopping times, as the defining events depend on the entire future of the random walk. Nevertheless, they are finite:

Lemma 13.7.6 *For any initial state $x \in \Gamma$, with P^x-probability one, the random variables η_k are all finite.*

Proof. It suffices to show that for each $m \in \mathbb{N}$ there is, with P^x-probability one, some $\vartheta_{m+k} > \vartheta_m$ such that $X_n \in C^+(X_{\vartheta_{m+k}})$ for all $n \geq \vartheta_{m+l}$. Define (possibly infinite) stopping times $\vartheta_m = M_0 = N_0 < M_1 < N_1 < M_2 < \cdots$ inductively by

$$M_{j+1} = \min\left\{n > N_j \;:\; X_n \notin C^+(X_{N_j})\right\},$$
$$N_{j+1} = \min\left\{\vartheta_{m+l} > M_{j+1}\right\},$$

with the usual convention that if any M_j is infinite then all subsequent $N_{j'}$ and $M_{j'}$ are also infinite. On the event $\left\{M_j < \infty\right\}$, the random variable N_j is almost surely finite, by Lemma 13.7.5 and the Markov property. Moreover, at each time $N_j < \infty$ the random walk is at a vertex $X_{N_j} = y$ whose cone type $\tau(y)$ is recurrent, and M_{j+1} is the first time after — if there is one — that the random walk exits the fat cone $C^+(y)$. Now as long as the random walk remains in $C^+(y)$, it behaves exactly as a translate of a random walk initiated at any other state $y' \in \Gamma$ with cone type τ_y^+; consequently, by the Markov property and Proposition 13.7.4, for each $j \in \mathbb{N}$,

$$\begin{aligned} P^x(\left\{M_{j+1} < \infty\right\} \cap \left\{M_j < \infty\right\}) &= P^x(\left\{M_{j+1} < \infty\right\} \cap \left\{N_j < \infty\right\}) \\ &\leq P^x\left\{M_j < \infty\right\}(1 - p_*), \end{aligned}$$

where $p_* = \min_g p(\tau^+(g))$ is the minimum of the no-exit probabilities (13.7.3) over all group elements g with recurrent cone types. Since there are only finitely many cone types, $p_* > 0$. Therefore, $P^x\left\{M_j < \infty\right\} \leq (1 - p_*)^j \to 0$ as $j \to \infty$, and so it follows that with probability one, $M_j = \infty$ for some $j \in \mathbb{N}$. □

Proof of Proposition 13.7.1. At each time η_k, the random walk is at a vertex x with recurrent cone type and has just completed a tour of the ball $B_{2\delta+1}(x)$. Subsequent to time η_k, the random walk remains forever in the fat cone $C^+(x)$. By 13.3.16, any geodesic ray γ with initial point 1 that converges to a boundary point in $\partial C^+(x)$ must pass through the ball $B_{2\delta+1}(x)$; consequently, the random walk must have visited γ during its tour of this ball. □

Additional Notes. Hyperbolic metric spaces and hyperbolic groups were introduced by Gromov [59], who is responsible for many of the results in Sections 13.1–13.4. There are a number of excellent surveys of the theory of hyperbolic groups; one of the very best is that of Ghys and de la Harpe [50]. The results of Section 13.5 are due to Kaimanovich [70]. Earlier, Ancona [3] had proved that for random walks with finitely supported step distributions, the full *Martin boundary* (see

Note 11.6.8) coincides with the Gromov boundary; this implies that for such random walks the Gromov boundary is also a Poisson boundary. In an even earlier article, C. Series [115] had proved Ancona's theorem for the special case where the underlying group is Fuchian. Important extensions of Kaimanovich's results to large classes of *weakly hyperbolic* groups have been obtained by Maher and Tiozzo [94]. Proposition 13.7.4 is taken from the article [63]of Haïssinsky, Mathieu, and Müller. This article also uses cone points to prove a central limit theorem for random walks on Fuchsian groups.

Chapter 14
Unbounded Harmonic Functions

14.1 Lipschitz Harmonic Functions

We proved in Chapter 10 that a symmetric random walk on a finitely generated group has nonconstant bounded harmonic functions if and only if it has positive Avez entropy. Any nearest-neighbor random walk on a group of subexponential growth must have Avez entropy 0, so for such random walks the only bounded harmonic functions are constant. Nevertheless, there are symmetric random walks with Avez entropy 0 that have *unbounded* harmonic functions: for instance, for the simple random walk on $\mathbb{Z}$ the function $h(x) = x$ is harmonic. So one might ask: for random walks with Avez entropy 0, are there always nonconstant harmonic functions?

Theorem 14.1.1 *For each infinite, finitely generated group* Γ*, every irreducible, symmetric random walk on* Γ *whose step distribution has finite support admits nonconstant, Lipschitz harmonic functions.*

This will be proved in Section 14.3. Here the term *Lipshitz* refers to the word metric on the group Γ: a function $f : \Gamma \to \mathbb{R}$ is Lipschitz with respect to the word metric induced by the symmetric generating set $\mathbb{A}$, with Lipschitz constant $C < \infty$, if

$$\sup_{x \in \Gamma} \max_{a \in A} |f(xa) - f(x)| \leq C. \tag{14.1.1}$$

The space of all *centered* Lipschitz harmonic functions $h : \Gamma \to \mathbb{R}$ (i.e., harmonic functions satisfying $h(1) = 0$, where 1 is the group identity) will be denoted by $\mathcal{H}_L$. This space is a real vector space, and it has a norm $\|\cdot\|_L$ defined by

$$\|h\|_L := \inf \left\{ C \in \mathbb{R} \, : \, \sup_{x \in \Gamma} \max_{a \in A} |f(xa) - f(x)| \leq C \right\}. \tag{14.1.2}$$

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_14

Although the norm $\|\cdot\|_L$ depends on the choice of the generating set $\mathbb{A}$, the space $\mathcal{H}_L$ does not (cf. inequality (1.2.4)). Consequently, there is no loss if generality in assuming that the step distribution μ of the random walk has support $\mathbb{A}$.

Exercise 14.1.2 The normed vector space $(\mathcal{H}_L, \|\cdot\|_L)$ is a Banach space, that is, it is complete in the metric induced by the norm (14.1.2).

Exercise 14.1.3 Let $h \in \mathcal{H}_L$ be a Lipschitz harmonic function. Show that for any $x \in \Gamma$ the functions $\mathcal{L}_x h$ and $\mathcal{L}_x^0 h$ defined by

$$\begin{aligned} \mathcal{L}_x h(y) &= h(x^{-1}y) \quad \text{and} \\ \mathcal{L}_x^0 h(y) &= h(x^{-1}y) - h(x^{-1}) \end{aligned} \tag{14.1.3}$$

are harmonic, and have the same Lipschitz norm as h. In addition, show that $\mathcal{L}^0$ is Γ-action on $\mathcal{H}_L$, that is,

$$\mathcal{L}^0_{x_1 x_2} = \mathcal{L}^0_{x_1} \circ \mathcal{L}^0_{x_2} \quad \text{for all} \;\; x_1, x_2 \in \Gamma. \tag{14.1.4}$$

Finally, show that if the step distribution μ is symmetric then the function

$$Jh(y) = h(y^{-1}) \tag{14.1.5}$$

is harmonic, with the same Lipschitz norm as h.

Thus, the mapping $x \mapsto \mathcal{L}_x^0$ is a *representation* of the group Γ in the group of linear isometries of the Banach space $\mathcal{H}_L$. This observation is especially useful when Γ is a group of polynomial growth.

Definition 14.1.4 A finitely generated group Γ has *polynomial growth* with respect to the (finite, symmetric) generating set $\mathbb{A}$ if there are constants $C, d < \infty$ such that the ball $\mathbb{B}_m$ of radius m in the Cayley graph satisfies

$$|\mathbb{B}_m| \leq Cm^d \quad \text{for infinitely many integers } m \geq 1. \tag{14.1.6}$$

Exercise 14.1.5 Suppose that Γ is a finitely generated group of polynomial growth, with growth exponent

$$d := \liminf_{m \to \infty} \frac{\log |\mathbb{B}_m|}{\log m} \tag{14.1.7}$$

relative to some finite, symmetric generating set $\mathbb{A}$.

(A) Show that Γ has polynomial growth rate relative to *any* finite, symmetric generating set, with the same growth exponent d.
(B) Show that if $\varphi : \Gamma \to \Gamma'$ is a surjective homomorphism then (i) Γ' is finitely generated, and (ii) Γ' has polynomial growth with growth exponent $\leq d$.

Theorem 14.1.6 (Colding & Minicozzi [26]; Kleiner [80]) *If Γ is an infinite, finitely generated group of polynomial growth then for any symmetric, nearest-neighbor random walk on Γ whose step distribution satisfies* $\min_{a \in \mathbb{A}} \mu(a) > 0$, *the space $\mathcal{H}_L$ of centered Lipschitz harmonic functions is finite-dimensional.*

The hypothesis on the step distribution is needed: for instance if $\Gamma = \mathbb{Z}^2$ and μ is the uniform distribution on $\{(1, 0), (-1, 0)\}$ then the space $\mathcal{H}_L$ is trivially infinite-dimensional (Exercise!). Theorem 14.1.6 will be proved in Section 14.5 below.

14.2 Harmonic Functions on Virtually Abelian Groups†

A group Γ is *virtually abelian* if it has an abelian subgroup H of finite index (which, by Exercise 1.2.11, must itself be finitely generated). According to the Structure Theorem for finitely generated abelian groups (see, for instance, Herstein [65], Theorem 4.5.1), H must be isomorphic to a direct sum $A \oplus \mathbb{Z}^d$ of a finite abelian group A and an integer lattice $\mathbb{Z}^d$, that is,

$$H \cong A \oplus \mathbb{Z}^d. \tag{14.2.1}$$

Therefore, a group is virtually abelian if and only if it has $\mathbb{Z}^d$ as a finite index subgroup, for some $d \geq 0$, called the *rank* of the group. In this section we will investigate the space of Lipschitz harmonic functions for a random walk on such a group.

Exercise 14.2.1 Let $\Gamma = \mathbb{Z}^d$ and let μ be any symmetric probability distribution on Γ with finite first moment

$$\sum_{x \in \mathbb{Z}^d} |x| \mu(\{x\}) < \infty.$$

(Here $|x|$ denotes word norm on $\mathbb{Z}^d$ with respect to the standard generating set, equivalently, the L^1-norm.)

(A) Show that each coordinate function $u_i(x) = x_i$ is μ-harmonic. Thus, $\dim(\mathcal{H}_L) \geq d$.
(B) Show that if the random walk with step distribution μ is irreducible, then every centered, Lipschitz harmonic function is a linear combination of the coordinate functions u_i. Therefore, $\dim(\mathcal{H}_L) = d$.

HINT: For each index $i \leq d$ the function $D_i h(x) := h(x + e_i) - h(x)$ is a *bounded* harmonic function. Now see Proposition 9.6.7.

Exercise 14.2.2 Let $\varphi : \Gamma \to \Gamma'$ be a surjective homomorphism of finitely generated groups, let μ be a probability distribution on Γ, and let μ' be the

pushforward of μ, that is, the probability distribution on Γ' defined by

$$\mu'(y) = \sum_{x\,:\,\varphi(x)=y} \mu(x). \tag{14.2.2}$$

(A) Show that if $u : \Gamma' \to \mathbb{R}$ is μ'-harmonic then its pullback $u \circ \varphi : \Gamma \to \mathbb{R}$ is μ-harmonic.
(B) Conclude that for any symmetric, irreducible random walk on a finitely generated abelian group of rank d, $\dim(\mathcal{H}_L) \geq d$. (Proposition 14.2.3 below will show that the dimension is exactly d.)

Next, let's consider virtually abelian groups Γ. Keep in mind that such a group Γ need not be abelian (cf. Exercise 1.5.2), nor will it necessarily be the case that a finite-index abelian subgroup H is normal.

Proposition 14.2.3 *Let Γ be a finitely generated group with finite-index subgroup $H \cong \mathbb{Z}^d$, and let $(X_n)_{n\geq 0}$ be an irreducible random walk on Γ whose step distribution μ is symmetric and has finite support.*

(A) If $u : \Gamma \to \mathbb{R}$ is any centered μ-harmonic function of at most linear growth, then the restriction of u to H is a linear combination of the coordinate functions on H. Furthermore, u is uniquely determined by its restriction to H.
(B) For each coordinate function $v_i : H \cong \mathbb{Z}^d \to \mathbb{Z}$ there is a unique Lipschitz harmonic function u whose restriction to H is v_i, and so

$$dim(\mathcal{H}_L) = d. \tag{14.2.3}$$

(C) For every $g \in H$ the (centered) translation operator $\mathcal{L}^0_g$ on $\mathcal{H}_L$ is the identity.

Assertion (B) has the following extension.

Proposition 14.2.4 *Let Γ be a finitely generated group with a finite-index subgroup H for which there exists a surjective homomorphism $\varphi : H \to \mathbb{Z}^d$. Let $(X_n)_{n\geq 0}$ be an irreducible random walk on Γ whose step distribution μ is symmetric and has finite support. Then for each coordinate function $v_i : \mathbb{Z}^d \to \mathbb{Z}$ the pullback $w_i := v_i \circ \varphi : H \to \mathbb{Z}$ extends to a Lipschitz harmonic function on Γ. Consequently,*

$$dim(\mathcal{H}_L) \geq d. \tag{14.2.4}$$

Proof of Proposition 14.2.3. **(A)** By hypothesis, Γ has a subgroup $H \cong \mathbb{Z}^d$ of finite index. Let $(X_n)_{n\geq 0}$ be a random walk on Γ with step distribution μ and increments $\xi_n := X_{n-1}^{-1} X_n$, and let

$$T := \min\{n \geq 1 \,:\, X_n \in H\} \tag{14.2.5}$$

be the first entrance time for H. By Lemma 7.4.5, $P^x\{T < \infty\} = 1$ for every $x \in \Gamma$, and the expectations $E^x T$ are uniformly bounded.

Since the step distribution has finite support, for any function $u : \Gamma \to \mathbb{R}$ the random variables $|u(X_n)|$ have finite expectations under P^x, for any $x \in \Gamma$. Consequently, if u is harmonic, then for any initial point $x \in \Gamma$ the sequence $(u(X_n))_{n\geq 0}$ is a martingale under P^x relative to the standard filtration (cf. Example 8.1.2), and so Doob's Optional Stopping Formula (Corollary 8.2.4) implies that

$$E^x u(X_{T\wedge n}) = u(x) \quad \text{for all } n \in \mathbb{N}. \tag{14.2.6}$$

We will deduce from this and the fact that the expectations $E^x T$ are finite that if u has at most linear growth then

$$E^x u(X_T) = u(x) \quad \text{for all } x \in \Gamma. \tag{14.2.7}$$

This will imply that u is uniquely determined by its values on H, and also that u is harmonic for the induced random walk $(X_{T_m})_{m\geq 0}$ on $H \cong \mathbb{Z}^d$ (see Proposition 7.4.3). Since the distribution of the induced random walk on H has finite first moment, by Lemma 7.4.9, it will then follow from Exercises 14.2.1 that the (centered) restriction of u to H is a finite linear combination of the coordinate functions.

Assume that u is harmonic and has at most linear growth; then there are constants $0 < B, C < \infty$ such that $|u(x)| \leq C|x| + B$ for all $x \in \Gamma$. Since the step distribution μ of the random walk has finite support, there exists $C' < \infty$ depending only on support(μ) such that $|X_n| \leq |X_0| + C'n$ for any $n \in \mathbb{N}$. Consequently, for any initial point $x \in \Gamma$, with $C_x := C' + |x|$,

$$|X_n| \leq C_x n \quad \text{for all } \; n \in \mathbb{N} \tag{14.2.8}$$

with P^x-probability 1. Thus,

$$\begin{aligned}
|E^x u(X_{T\wedge m}) - E^x u(X_T)| =& |E^x u(X_{T\wedge m})\mathbf{1}\{T > m\} - E^x u(X_T)\mathbf{1}\{T > m\}| \\
\leq& C E^x |X_{T\wedge m}|\mathbf{1}\{T > m\} \\
&+ C E^x |X_T|\mathbf{1}\{T > m\} \\
&+ 2BP\{T > m\} \\
\leq& 2CC_x E^x T\mathbf{1}\{T > m\} + 2BP^x\{T > m\}.
\end{aligned}$$

By Lemma 7.4.5, $E^x T < \infty$, so both terms in the preceding bound converge to 0 as $m \to \infty$, by the dominated convergence theorem. Hence, (14.2.7) follows from (14.2.6). This proves the assertion (A). □

Proof of Proposition 14.2.4. Let $(v_i)_{i \leq d}$ be the coordinate functions on $\mathbb{Z}^d$, and let $w_i = v_i \circ \varphi$ be the pullback of v_i to H. For each index i define $u_i : \Gamma \to \mathbb{R}$ by

$$u_i(x) = E^x w_i(X_\tau) \quad \text{where} \quad \tau := \min\{n \geq 0 : X_n \in H\}. \tag{14.2.9}$$

We will show that u_i is well-defined, harmonic, and Lipschitz on Γ. Clearly, $u_i = w_i$ on H, because if $x \in H$ then $P^x\{\tau = 0\} = 1$; consequently, the functions u_i are linearly independent. Equation (14.2.7) implies that any harmonic function of linear growth on Γ is uniquely determined by its restriction to H, so the uniqueness part of assertion (B) in Proposition 14.2.3 will follow.

First, let's check that the functions $w_i = v_i \circ \varphi$ are Lipshitz on H. Recall (Exercise 1.2.11) that any finite-index subgroup of a finitely generated group is finitely generated, so H has a finite generating set $\{h_j\}_{j \leq J}$. Since $\varphi : H \to \mathbb{Z}^d$ is a homomorphism, for any word $h_{j_1} h_{j_2} \cdots h_{j_n} h_{j_{n+1}}$ in the generators h_j,

$$\begin{aligned} &|v_i \circ \varphi(h_{j_1} h_{j_2} \cdots h_{j_{n+1}}) - v_i \circ \varphi(h_{j_1} h_{j_2} \cdots h_{j_{n+1}})| \\ &= \left| v_i\left(\sum_{j=1}^{n+1} \varphi(h_{j_j})\right) - v_i\left(\sum_{j=1}^{n} \varphi(h_{j_j})\right) \right| \\ &\leq \max_{j \leq J} |\varphi(h_j)|. \end{aligned} \tag{14.2.10}$$

Next, we will show that the expectation (14.2.9) is well-defined for $x \notin H$. If $x \notin H$ then under P^x the random variable τ coincides with the first-entry time T (defined by equation (14.2.5)), so by Lemma 7.4.3 the expectations $E^x \tau$ are uniformly bounded. Since the step distribution has finite support,

$$d(X_0, X_\tau) \leq C_1 \tau, \tag{14.2.11}$$

where d is the word metric on Γ and $C_1 = \max\{|g| : \mu(g) > 0\}$. By (14.2) the functions $w_i = v_i \circ \varphi$ satisfy $|w_i(y)| \leq C|y|$ for a suitable constant $C < \infty$; consequently, by (14.2.11),

$$E^x |w_i(X_\tau)| \leq C E^x |X_\tau| \leq C|x| + C C_1 E^x \tau.$$

Thus, not only is u_i well-defined, but it has at most linear growth on Γ.

To prove that the functions u_i are harmonic, we must verify that for each index i the mean value property

$$u_i(z) = \sum_{y \in \Gamma} p(z, y) u_i(y) \tag{14.2.12}$$

holds at all points $z \in \Gamma$. For points $z \notin H$ this can be done by essentially the same argument (condition on the first step!) as in Proposition 6.2.9 (Exercise: Check this.)

Thus, our task is to establish the identity (14.2.12) for points $z \in H$. For any initial point $y \notin H$ the random variables τ and T are P^y-almost surely equal, and so $u_i(y) := E^y w_i(X_\tau) = E^y w_i(X_T)$; hence, by the Markov property, if $z \in H$ then

$$\begin{aligned}\sum_{y\in\Gamma} p(z,y)u_i(y) &= \sum_{y\in\Gamma\setminus H} p(z,y)E^y w_i(X_T) + \sum_{y\in H} p(z,y)w_i(y)\\ &= E^z w_i(X_T) = E^z u_i(X_T).\end{aligned}$$

Therefore, to prove that the mean value property (14.2.12) holds at points $z \in H$, it will suffice to show that for all $z \in H$,

$$E^z u_i(X_T) = u_i(z).$$

This is equivalent to showing that the restriction $w_i = v_i \circ \varphi$ of u_i to the subgroup H is harmonic for the *induced* random walk (cf. Proposition 7.4.3). But this follows by Exercises 14.2.1 and 14.2.2. In particular, Lemma 7.4.7 implies that the step distribution of the induced random walk is symmetric, and Lemma 7.4.9 shows that it has finite first moment, so Exercise 14.2.1 implies that v_i is harmonic for the pushforward $(\varphi(X_{T_m}))_{m\geq 0}$ of the induced random walk; consequently, by Exercise 14.2.2, w_i is harmonic for the induced random walk.

It remains to show that the functions u_i are Lipschitz. Because the subgroup H has finite index in Γ, there exists $C_2 < \infty$ such that the distance from any $x \in \Gamma$ to H is $\leq C_2$. This is because every $x \in \Gamma$ is in one of finitely many right cosets Hx_i; if $x = yx_j$ for some $y \in H$ then $d(x, H) = d(x_j, H) \leq \max_i d(x_i, H) < \infty$. It follows that if $x, y \in \Gamma$ are nearest neighbors (that is, $d(x, y) = 1$), then there is an element $z \in H$ such that both x and y are within distance C_2+1 of z. Consequently, since $w_i = v_i \circ \varphi$ is Lipschitz on H, with Lipschitz constant $C_0 := \max_j |\varphi(h_j)|$,

$$\begin{aligned}|E^x w_i(X_\tau) - w_i(z)| &\leq C_0(C_2+1) + C_0 E^x d(X_\tau, x) \quad \text{and}\\ |E^y w_i(X_\tau) - w_i(z)| &\leq C_0(C_2+1) + C_0 E^y d(X_\tau, y).\end{aligned}$$

Therefore, by the triangle inequality and (14.2.11), if x, y are nearest neighbors then

$$|E^x w_i(X_\tau) - E^y w_i(X_\tau)| \leq 2C_0(C_2+1) + C_1 E^x \tau + C_1 E^y \tau.$$

But by Lemma 7.4.1, the expectations $E^x\tau$ and $E^y\tau$ are uniformly bounded, so this proves that the function $x \mapsto E^x w_i(X_\tau)$ is Lipschitz. □

Proof of Proposition 14.2.3. **(C)** This is almost trivial: since $H \cong \mathbb{Z}^d$, the elements $g \in H$ act as linear translations on $\mathbb{Z}^d$, and so for any $g \in H$ and any coordinate function $v_i : H \to \mathbb{Z}$,

$$v_i(gx) = v_i(g) + v_i(x) \quad \text{for all } x \in H.$$

Thus, $\mathcal{L}_g^0(v_i) = v_i$. Since the coordinate functions span $\mathcal{H}_L$, by assertion (B) of the proposition, it follows that $\mathcal{L}_g^0$ is the identity. □

In the remainder of this section, we will investigate the nature of *polynomial growth* harmonic functions on the integer lattices $\mathbb{Z}^d$.

Exercise 14.2.5 † Let μ be a symmetric probability distribution on $\mathbb{Z}$ with support $\mathbb{A}_K := \{k \in \mathbb{Z} : |k| \leq K\}$ such that $\mu(k) > 0$ for every $k \in \mathbb{A}_K$, and let $\mathcal{H}$ be the real vector space of all μ-harmonic functions on $\mathbb{Z}$.

(A) Show that $\dim(\mathcal{H}) = 2K$.
(B) Show that $\exists\, R < \infty$ such that $\forall u \in \mathcal{H}$,

$$\limsup_{m\to\infty} \max(|u(m)|, |u(-m)|)^{1/m} \leq R.$$

Conclude that the power series

$$U_+(z) = \sum_{m=0}^{\infty} u(m)z^m \quad \text{and} \quad U_-(z) = \sum_{m=0}^{\infty} u(-m)z^m$$

have radii of convergence at least $1/R$.
(C) Show that for any $u \in \mathcal{H}$ there exist polynomials $G_+(z), G_-(z)$ of degree $\leq 2K$ such that

$$U_+(z) = \frac{G_+(z)}{z^K - z^K F(z)} \quad \text{and} \quad U_-(z) = \frac{G_-(z)}{z^K - z^K F(z)} \quad \text{where}$$
$$F(z) = \sum_{k\in\mathbb{A}_K} \mu(k)z^k.$$

(D) Show that $z = 1$ is a root of the polynomial $z^K - z^K F(z)$ of (exact) order 2.
(E) Use partial fraction decomposition together with the results of (C) and (D) to conclude that if $u \in \mathcal{H}$ is not a linear function then it has exponential growth at either $-\infty$ or $+\infty$.

Exercise 14.2.1 shows that in any dimension d the linear functions are harmonic for any symmetric random walk whose step distribution has finite support, and Exercise 14.2.5 shows that the *only* harmonic functions for the simple random walk in $\mathbb{Z}^1$ are the linear functions $u(x) = Ax + B$. Are there other harmonic functions in higher dimensions? A bit of experimentation will reveal that there are indeed others: for example, the quadratic functions $u(x) = x_1^2 - x_2^2$ and $v(x) = x_1x_2$ are harmonic for the simple random walk on $\mathbb{Z}^2$, as is the quartic polynomial $w(x) = x_1^4 - x_2^4 + Cx_1^2x_2^2$ for the right choice of C (exercise!), and so on. Are there harmonic functions of polynomial growth that are not polynomials? The following exercise shows that the answer is no.

WARNING: This exercise will require some knowledge of Fourier analysis and the theory of Schwartz distributions, as might be found, for instance, in [114], Chapter 1. Here is a brief review of the key facts. Denote by $\mathbb{T} = [-\pi, \pi]^n$ the standard n-dimensional torus (with opposite faces identified), let $\mathcal{P} = C^\infty(\mathbb{T})$ be the space of infinitely differentiable complex-valued functions on $\mathbb{T}$, equipped with the standard Fréchet topology, and let $\mathcal{P}'$ be the space of Schwartz distributions on $\mathbb{T}$. (See [114], Section 2.3.) Every function $f \in \mathcal{P}$ has a *Fourier series*

$$f(\boldsymbol{\theta}) = \sum_{\mathbf{m}\in\mathbb{Z}^n} \hat{f}(\mathbf{m})e^{i\mathbf{m}\cdot\boldsymbol{\theta}},$$

whose coefficients $\hat{f}(\mathbf{m})$ decay faster than any polynomial in the entries $\mathbf{m}$, so for any function $u : \mathbb{Z}^d \to \mathbb{R}$ of at most polynomial growth there is a Schwartz distribution U such that

$$\langle U, f\rangle = \sum_{\mathbf{m}\in\mathbb{Z}^n} \hat{f}(\mathbf{m})u(\mathbf{m}).$$

The function u is called the *Fourier transform* of the Schwartz distribution U.

Exercise 14.2.6 † Let μ be a symmetric probability distribution on a finite, symmetric generating set $\mathbb{A}$ of $\mathbb{Z}^n$, and let $\hat{\mu}$ be the Fourier transform of μ, defined by

$$\hat{\mu}(\boldsymbol{\theta}) = \sum_{\mathbf{m}\in\mathbb{Z}^n} \mu(\mathbf{m})e^{i\mathbf{m}\cdot\boldsymbol{\theta}} = \sum_{\mathbf{m}\in\mathbb{A}} \mu(\mathbf{m})e^{i\mathbf{m}\cdot\boldsymbol{\theta}}.$$

(A) Show that $\hat{\mu}$ is real-valued and satisfies $-1 < \hat{\mu}(\boldsymbol{\theta}) < 1$ for all $\boldsymbol{\theta} \in \mathbb{T} \setminus \{\mathbf{0}\}$.
(B) Let u be a μ-harmonic function that has at most polynomial growth at infinity, and let U be the Schwartz distribution with Fourier transform u. Show that for every $f \in \mathcal{P}$,

$$\langle U, f\rangle = \left\langle U, (1-\hat{\mu})f\right\rangle.$$

(C) Deduce from this that the distribution U has support $\{\mathbf{0}\}$.
(D) Deduce that $U = \sum_{\alpha\in F} a_\alpha D^\alpha \delta_{\mathbf{0}}$ for some finite set F of multi-indices α. (This follows from [114], Theorem 2.4.5, and (D); alternatively, see Section 6.3 of [120].)
(E) Conclude that $u(x)$ is a polynomial in the coordinates x_i.

Remark 14.2.7 The argument presented in Exercise 14.2.6 also applies, with little change, to show that the polynomial-growth harmonic functions for the (standard) Laplace operator $\Delta = \sum_{i=1}^n \partial_i^2$ are all polynomials. See Appendix B of [88] for a different approach.

14.3 Existence of Nonconstant Harmonic Functions

Kesten's theorem (Chapter 5) implies that every irreducible, symmetric random walk on a nonamenable group has positive Avez entropy, and therefore (cf. Corollary 10.3.4) admits non-constant, bounded harmonic functions. Thus, the existence of nonconstant Lipschitz harmonic functions is only at issue for *amenable* groups, so to prove Theorem 14.1.1 we may assume that the group Γ is amenable.

Assumption 14.3.1 *Assume for the remainder of this section and in Section 14.5 that Γ is* amenable *and that μ is the step distribution of a nearest-neighbor random walk $(X_n)_{n\geq 0}$ on Γ, with Markov operator $\mathbb{M}$ and Dirichlet form $\mathcal{D}(\cdot,\cdot)$. Assume also that $\mu(a) > 0$ for every generator a, and (for ease of exposition) that $\mu(1) = 0$.*

The assumptions on the step distribution μ do not entail any loss of generality, for reasons we have seen earlier: (i) lazy variants of a random walk have the same harmonic functions, and (ii) for an irreducible, nearest-neighbor random walk, the generating set can always be adjusted so that every generator is given positive probability by the step distribution.

The proof of Theorem 14.1.1 will use the notion of *convolution* $f * g$ of two functions $f, g : \Gamma \to \mathbb{R}$. This is defined as follows:

$$f * g(x) = \sum_{z\in\Gamma} f(z)g(z^{-1}x) = \sum_{z\in\Gamma} f(z)\mathcal{L}_z g(x), \tag{14.3.1}$$

provided the series is convergent. (Here $\mathcal{L}_z$ is the left translation operator defined in Exercise 14.1.3.) By the Cauchy-Schwarz inequality, the series defining $f * g$ is convergent whenever $f, g \in L^2(\Gamma)$.

Lemma 14.3.2

$$\inf_{g\in L^2(\Gamma)\,:\,\mathcal{D}(g,g)=1} \|(I-\mathbb{M})g\|_2^2 = 0 \tag{14.3.2}$$

Note: The infimum is not attained, because if $\|g - \mathbb{M}g\|_2 = 0$ for some function $g \in L^2(\Gamma)$ then g would be harmonic, with limit $\lim_{|x|\to\infty} g(x) = 0$. By the Maximum Principle for harmonic functions, any such g would necessarily be identically 0.

Proof of Theorem 14.1.1. Before proving Lemma 14.3.2 we will show how it implies Theorem 14.1.1. The central idea is this: (a) if $\|(I - \mathbb{M})g\|_2$ is small, then g must be nearly harmonic, because harmonicity is equivalent to $(I - \mathbb{M})g = 0$; and (b) if $\mathcal{D}(g, g) = 1$, then at least some of the nearest-neighbor differences $|g(x) - g(y)|$ must be large, so g cannot be approximately constant. The strategy, then, will be to find a pointwise convergent sequence of such approximately harmonic, nonconstant functions.

By definition of the Dirichlet form,

$$\begin{aligned}\mathcal{D}(g,g) &= \frac{1}{2}\sum_{x\in\Gamma}\sum_{a\in A}(g(xa)-g(x))^2\mu(a)\\ &= \frac{1}{2}\sum_{x\in\Gamma}\sum_{a\in A}(g*\delta_{a^{-1}}(x)-g(x))^2\mu(a)\\ &= \frac{1}{2}\sum_{a\in A}\|g-g*\delta_{a^{-1}}\|_2^2\,\mu(a). \qquad (14.3.3)\end{aligned}$$

Here δ_y is the function that takes the value 1 at y and 0 elsewhere, and $*$ denotes convolution.

Equation (14.3.3) shows that if g has Dirichlet energy $\mathcal{D}(g,g)=1$ then (i) for every generator $a\in\mathbb{A}$ the function $g-g*\delta_a$ has squared L^2-norm bounded by $2/\mu(a)$; and (ii) for *some* generator a, the function $g-g*\delta_a$ has squared norm at least 2. Thus, by Lemma 14.3.2, there exist a generator $a\in A$ and a sequence $g_n\in L^2(\Gamma)$ such that

$$\begin{gathered}\mathcal{D}(g_n,g_n)=1,\\ 2\le\|g_n-g_n*\delta_a\|_2^2\le 2/\mu(a),\quad\text{and}\\ \lim_{n\to\infty}\|(I-\mathbb{M})g_n\|_2^2=0.\end{gathered}$$

Define $h_n:\Gamma\to\mathbb{R}$ by $h_n(x)=g_n(x^{-1})-g_n*\delta_a(x^{-1})$, and consider the convolution h_n*g_n. Since both h_n and g_n are elements of $L^2(\Gamma)$, the convolution h_n*g_n is well-defined, and is a linear combination of left translates of g_n, in particular,

$$h_n*g_n=\sum_{z\in\Gamma}h_n(z)\mathcal{L}_z g_n.$$

The Markov operator $\mathbb{M}$ commutes with left translations $\mathcal{L}_z$ (as you should check); hence,

$$(I-\mathbb{M})h_n*g_n=\sum_{z\in\Gamma}h_n(z)\mathcal{L}_z((I-\mathbb{M})g_n).$$

It follows, by the Cauchy-Schwarz inequality, that

$$\|(I-\mathbb{M})(h_n*g_n)\|_2\le\|h_n\|_2\,\|(I-\mathbb{M})g_n\|_2\le\sqrt{\frac{2}{\mu(a)}}\,\|(I-\mathbb{M})g_n\|_2\longrightarrow 0.$$

Cauchy-Schwarz and the associativity of convolution also imply that for every generator $b \in \mathbb{A}$ and every group element $x \in \Gamma$

$$\begin{aligned} |h_n * g_n(x) - h_n * g_n(xb)| &= |h_n * (g_n - g_n * \delta_b)(x)| \\ &\leq \|h_n\|_2 \, \|g_n - g_n * \delta_b\|_2 \\ &\leq 2/\sqrt{\mu(a)\mu(b)}. \end{aligned}$$

Consequently, the function $h_n * g_n$ is Lipschitz, with Lipschitz norm no larger than $\max_{b \in A} 2/\sqrt{\mu(a)\mu(b)}$. On the other hand, the Lipschitz norms are at least 2, because for each n,

$$\begin{aligned} h_n * g_n(1) - h_n * g_n(a) &= h_n * (g_n - g_n * \delta_a)(1) \\ &= \|g_n - g_n * \delta_a\|_2^2 \geq 2. \end{aligned}$$

Now let $f_n : \Gamma \to \mathbb{R}$ be the function obtained by centering $h_n * g_n$, that is,

$$f_n(x) = h_n * g_n(x) - h_n * g_n(1).$$

Centering a function does not affect the Lipschitz norm, so each f_n has Lipschitz norm bounded by $\max_{a,b \in A} 2/\sqrt{\mu(a)\mu(b)}$. Hence, the functions f_n are uniformly bounded on each ball $\mathbb{B}_k$ in Γ, and so by the Bolzano-Weierstrass theorem there is a pointwise-convergent subsequence f_{n_m}, with limit

$$f(x) := \lim_{m \to \infty} f_{n_m}(x).$$

The limit function f is 0 at the group identity, since all of the functions f_m are centered, but $|f(a)| \geq 2$, so f is not constant. Finally, f must be harmonic, because the limit relation

$$\lim_{n \to \infty} \|(I - \mathbb{M}) f_n\|_2 = \lim_{n \to \infty} \|(I - \mathbb{M})(h_n * g_n)\|_2 = 0$$

implies that $\lim_{n \to \infty} (I - \mathbb{M}) f_n(x) = 0$ for every $x \in \Gamma$. □

It remains to prove Lemma 14.3.2. There are two distinct cases, depending on whether the group Γ is recurrent or transient.

Proof. ***(Transient Groups)*** Recall (cf. Proposition 5.4.2) that for any function $f \in L^2(\Gamma)$,

$$\mathcal{D}(f, f) = \langle f, (I - \mathbb{M} f) \rangle .$$

Therefore, to prove the lemma, it suffices to show that for any $\varepsilon > 0$ there is a nonzero function $g \in L^2(\Gamma)$ such that

$$\frac{\|(I - \mathbb{M})g\|_2^2}{\langle g, (I - \mathbb{M})g \rangle} < \varepsilon. \tag{14.3.4}$$

By hypothesis, the group Γ is amenable, so by Exercise 5.1.3 it has isoperimetric constant 0 for every generating set. Fix a (symmetric) generating set $\mathbb{A}$; then for any $m \in \mathbb{N}$ the set $\cup_{i=1}^{m} \mathbb{A}^i$ is also a generating set; consequently, for any $m \in \mathbb{N}$ the group contains an increasing sequence $B_1 \subset B_2 \subset B_3 \subset \cdots$ of finite subsets such that

$$\lim_{k \to \infty} |\{x \in B_k \, : \, d(x, \partial B_k) \le m\}|/|B_k| = 0, \tag{14.3.5}$$

where $|\cdot|$ and d are the word norm and metric for the generating set $\mathbb{A}$. Fix $k \in \mathbb{N}$, let $f = \mathbf{1}_{B_k}$ be the indicator function of the set B_k, and define $g = \sum_{i=0}^{n-1} \mathbb{M}^i f$, where n is a large integer. We will show that for any $\varepsilon > 0$ there exists positive integers m, k, n such that (14.3.4) holds.

The rationale for trying a function g of the form $g = \sum_{i=0}^{n-1} \mathbb{M}^i f$ is that application of the operator $I - \mathbb{M}$ produces a telescoping sum: hence,

$$(I - \mathbb{M})g = f - \mathbb{M}^n f. \tag{14.3.6}$$

Since the random walk is transient, Pólya's criterion implies that the n-step transition probabilities $p_n(x, y)$ are summable for any fixed pair $x, y \in \Gamma$, and so in particular $\lim_{n\to\infty} p_n(x, y) = 0$. This implies that for any function $u : \Gamma \to \mathbb{R}$ with finite support, $\lim_{n\to\infty} \mathbb{M}^n u = 0$ pointwise. Consequently, by (14.3.6) and the symmetry of the Markov operator $\mathbb{M}$,

$$\begin{aligned}
\|(I - \mathbb{M})g\|_2^2 &= \left\| f - \mathbb{M}^n f \right\|_2^2 \\
&= \|f\|_2^2 - 2\left\langle f, \mathbb{M}^n f \right\rangle + \left\|\mathbb{M}^n f\right\|_2^2 \\
&= \|f\|_2^2 - 2\left\langle f, \mathbb{M}^n f \right\rangle + \left\langle f, \mathbb{M}^{2n} f \right\rangle \\
&= \|f\|_2^2 + o(1) \\
&= |B_k| + o(1) \quad \text{as } n \to \infty.
\end{aligned} \tag{14.3.7}$$

Next, let's estimate the denominator $\langle g, (I - \mathbb{M})g \rangle$ in (14.3.4). Recall that the transition probabilities of a transient random walk are *uniformly* summable (cf. Exercise 1.3.6), that is, for any $\varepsilon > 0$ there exists $n_\varepsilon < \infty$ such that for every pair $x, y \in \Gamma$,

$$\sum_{n \ge n_\varepsilon} p_n(x, y) < \varepsilon/|B_k|.$$

Thus, for any $n \ge n_\varepsilon$,

$$
\begin{aligned}
\langle g, (I - \mathbb{M})g \rangle &= \left\langle \sum_{i=0}^{n-1} \mathbb{M}^i f, f - \mathbb{M}^n f \right\rangle \\
&= \sum_{i=0}^{n-1} \sum_{x,y \in B_k} (p_i(x, y) - p_{n+i}(x, y)) \\
&\geq \sum_{i=0}^{n-1} \sum_{x,y \in B_k} p_i(x, y) - \varepsilon |B_k|. \qquad (14.3.8)
\end{aligned}
$$

Since the random walk is nearest-neighbor, if its initial point is an element of B_k at distance greater than m from ∂B_k then it must remain in B_k for its first m steps. Thus, by relation (14.3.5), for all k sufficiently large,

$$
\sum_{i=0}^{m} \sum_{x,y \in B_k} p_i(x, y) = \sum_{x \in B_k} \sum_{i=0}^{m} E^x \mathbf{1}_{B_k}(X_i) \geq m|B_k|(1 - \varepsilon). \qquad (14.3.9)
$$

Consequently, by (14.3.7), (14.3.8), and (14.3.9), for all sufficiently large n,

$$
\frac{\langle g, (I - \mathbb{M})g \rangle}{\|(I - \mathbb{M})g\|_2^2} \geq m(1 - \varepsilon) - \varepsilon.
$$

Since m can be chosen arbitrarily large and ε arbitrarily small, the lemma follows. □

Proof. (Recurrent Groups) The proof for transient groups fails for recurrent groups, because the asymptotic evaluation of the inner product $\langle g, (I - \mathbb{M})g \rangle$ relied critically on the summability of the n-step transition probabilities. By Pólya's criterion, for any recurrent random walk,

$$
\sum_{n=0}^{\infty} p_n(x, x) = \infty,
$$

and so the tail-sums do not converge to 0. Nevertheless, because the group Γ is infinite, any irreducible, symmetric, nearest-neighbor random walk is *null recurrent*, that is, $\lim_{n\to\infty} p_n(x, y) = 0$ for any $x, y \in \Gamma$. (This follows by Exercise 9.5.3 and Theorem 9.5.5: the latter implies that if the random walk has no nonconstant, *bounded* harmonic functions then it is weakly ergodic, and hence satisfies the hypothesis of Exercise 9.5.3).

Set $g = g_n = \sum_{i=0}^{n-1} \mathbb{M}^i \delta_1$, where δ_1 is the indicator of the set $\{1\}$ containing only the group identity. As in the transient case, $(I - \mathbb{M})g_n = \delta_1 - \mathbb{M}^n \delta_1$, so null recurrence implies that

$$\|(I - \mathbb{M})g_n\|_2^2 = \|\delta_1\|_2^2 + o(1) = 1 + o(1).$$

(Exercise: Verify this.) Therefore, to complete the proof it will suffice to show that

$$\lim_{n\to\infty} \langle g_n, (I - \mathbb{M})g_n \rangle = \infty. \tag{14.3.10}$$

By the symmetry of the Markov operator,

$$\begin{aligned}\langle g_n, (I - \mathbb{M})g_n \rangle &= \sum_{i=0}^{n-1} \left\langle \mathbb{M}^i \delta_1, \delta_1 - \mathbb{M}^n \delta_1 \right\rangle \\ &= \sum_{i=0}^{n-1} p_i(1, 1) - \sum_{i=0}^{n-1} p_{n+i}(1, 1).\end{aligned}$$

Recurrence ensures that the first sum becomes large as n grows; but unlike in the transient case, the second sum might also become large. For this reason, we fix a large number C and partition the range of summation as $[0, n-1] = [0, 2m-1] \cup [2m, n-1]$, where $m = m(C)$ is chosen sufficiently large that $\sum_{i=0}^{2m-1} p_i(1, 1) \geq C$. Null recurrence then implies that

$$\lim_{n\to\infty} \sum_{i=0}^{2m-1} p_i(1, 1) - \sum_{i=0}^{2m-1} p_{n+i}(1, 1) = \sum_{i=0}^{2m-1} p_i(1, 1) \geq C.$$

Consequently, to prove that $\lim_{n\to\infty} \langle g_n, (I - \mathbb{M})g_n \rangle = \infty$ it is enough to show that for any fixed $m \geq 1$,

$$\liminf_{n\to\infty} \left(\sum_{i=2m}^{n-1} p_i(1, 1) - \sum_{i=2m}^{n-1} p_{n+i}(1, 1) \right) \geq 0. \tag{14.3.11}$$

By null recurrence, $p_{n\pm i}(1, 1) \to 0$ as $n \to \infty$ uniformly for $0 \leq i \leq 2m$; thus, to prove (14.3.11) it is enough to show that

$$\sum_{i=2m}^{n-1} p_i(1, 1) - \sum_{i=2m}^{n-1} p_{n-2m+i}(1, 1) \geq 0.$$

But by the same telescoping-sum calculation as above, if $h_n = \sum_{i=0}^{n-2m-1} \mathbb{M}^i (\mathbb{M}^m \delta_1)$ then

$$\langle h_n, (I - \mathbb{M})h_n \rangle = \sum_{i=0}^{n-2m-1} \left(\left\langle \mathbb{M}^i \mathbb{M}^m \delta_1, \mathbb{M}^m \delta_1 \right\rangle - \left\langle \mathbb{M}^{n-2m+i} \mathbb{M}^m \delta_1, \mathbb{M}^m \delta_1 \right\rangle \right)$$

$$= \sum_{i=2m}^{n-1} (p_i(1,1) - p_{n-2m+i}(1,1)).$$

This is evidently nonnegative, because $\langle h_n, (I - \mathbb{M})h_n \rangle = \mathcal{D}(h_n, h_n)$, where $\mathcal{D}$ is the Dirichlet form (cf. Proposition 5.4.2). □

14.4 The Poincaré and Cacciopoli Inequalities

The proof of the Colding-Minicozzi-Kleiner theorem will require two inequalities relating the natural inner product on $L^2(\Gamma)$ to the Dirichlet form on $L^2(\Gamma)$. These inequalities do not require that the group be of polynomial growth — they are valid generally for symmetric, nearest neighbor random walks on arbitrary finitely generated groups.

Assume throughout this section that μ is the step distribution of a symmetric, nearest neighbor random walk on a finitely generated group Γ. Recall that the Dirichlet form $\mathcal{D}_U$ for a subset $U \subset \Gamma$ of the ambient group Γ is defined by equation (7.1.1):

$$\mathcal{D}_U(g, g) = \sum_{\{x,y\} \in \mathcal{E}_U} p(x, y)(g(y) - g(x))^2,$$

where $p(x, y) = \mu(x^{-1}y)$ are the one-step transition probabilities and $\mathcal{E}_U$ is the set of edges in the Cayley graph both of whose endpoints are in U. For any $z \in \Gamma$ and any $R \in \mathbb{N}$, denote by $\mathbb{B}(z, R) = \mathbb{B}_R(z)$ the ball of radius R centered at z in the Cayley graph, and abbreviate

$$\mathbb{B}_R = \mathbb{B}_R(1) \quad \text{and} \quad |\mathbb{B}_R| = \text{card}(\mathbb{B}_R).$$

Proposition 14.4.1 (Poincaré Inequality) *If* $\kappa := \min_{a \in \mathbb{A}} \mu(a) > 0$ *then for every function* $f : \Gamma \to \mathbb{R}$ *with mean value* 0 *on the ball* $\mathbb{B}(z, R)$*,*

$$\|f \mathbf{1}_{\mathbb{B}(z,R)}\|_2^2 \le 2R^2 \kappa^{-1} \frac{|\mathbb{B}_{2R}|}{|\mathbb{B}_R|} \mathcal{D}_{\mathbb{B}(z,3R)}(f, f). \tag{14.4.1}$$

Proposition 14.4.2 (Cacciopoli Inequality) *For every* harmonic *function h, every* $z \in \Gamma$*, and every* $R = 1, 2, \cdots$*,*

$$\mathcal{D}_{\mathbb{B}(z,R)}(h, h) \le R^{-2} \|h \mathbf{1}_{\mathbb{B}(z,3R)}\|_2^2. \tag{14.4.2}$$

Proof of Proposition 14.4.1. Since both the natural inner product on $L^2(\Gamma)$ and the Dirichlet form(s) are invariant by left translations, we may assume without loss of

generality that $z = 1$. For brevity, we use the shorthand notation $\mathbb{B}_R = \mathbb{B}(1, R)$ for the ball of radius R centered at the group identity.

If f has mean value 0 on the ball $\mathbb{B}_R$ then

$$\begin{aligned}\sum_{x\in\mathbb{B}_R}\sum_{y\in\mathbb{B}_R}|f(x)-f(y)|^2 &= \sum_{x\in\mathbb{B}_R}\sum_{y\in\mathbb{B}_R}(f(x)^2+f(y)^2-2f(x)f(y))\\ &= \sum_{x\in\mathbb{B}_R}\sum_{y\in\mathbb{B}_R}(f(x)^2+f(y)^2)\\ &= 2|\mathbb{B}_R|\,\left\|f\mathbf{1}_{\mathbb{B}_R}\right\|_2^2.\end{aligned}$$

Thus, to prove the Poincaré inequality (14.4.1) it will suffice to show that

$$\sum_{x\in\mathbb{B}_R}\sum_{y\in\mathbb{B}_R}|f(x)-f(y)|^2 \le 4\kappa^{-1}R^2|\mathbb{B}_{2R}|\mathcal{D}_{\mathbb{B}_{3R}}(f,f). \tag{14.4.3}$$

For any pair of points $x, y \in \mathbb{B}_R$ there is at least one (and in general more than one) path of length $\le 2R$ in the Cayley graph connecting x and y. Any such path can be represented as the translate $x\gamma$ of a path $\gamma = g_0g_1g_2\cdots g_n$ of length $n \le 2R$ beginning at the group identity $g_0 = 1$; for this path, the difference $f(y) - f(x)$ is the sum of the differences $f(xg_i) - f(xg_{i-1})$. Consequently, by the Cauchy-Schwarz inequality, if $\kappa = \min_{a\in A} p(1, a)$ then

$$\begin{aligned}(f(y)-f(x))^2 &\le n\sum_{i=1}^{n}(f(xg_i)-f(xg_{i-1}))^2\\ &\le 2R\kappa^{-1}\sum_{i=1}^{n}(f(xg_i)-f(xg_{i-1}))^2p(xg_{i-1},xg_i)\end{aligned}$$

For each $g \in \Gamma$ let $\gamma_g = x_0x_1\cdots x_{|g|}$ be a shortest path in the Cayley graph from the group identity 1 to g. The translate $x\gamma_g$ has endpoint xg, and so the set of all translates $x\gamma_g$, where g ranges over the ball $\mathbb{B}_{2R}$, will contain a path connecting x to $y \in \mathbb{B}_R$, for every $y \in \mathbb{B}_R$. For any fixed $g \in \mathbb{B}_{2R}$ the path γ_g has length at most $2R$; hence, for any edge e of the Cayley graph there are at most $2R$ points $x \in \Gamma$ such that the edge e occurs in the translated path $x\gamma_g$. Any such edge must have both its endpoints in the ball $\mathbb{B}_{3R}$, and so

$$\begin{aligned}&\sum_{x\in\mathbb{B}_R}\sum_{y\in\mathbb{B}_R}(f(y)-f(x))^2\\ &\qquad\le \sum_{x\in\mathbb{B}_R}\sum_{g\in\mathbb{B}_{2R}}(f(xg)-f(x))^2\end{aligned}$$

$$\leq \sum_{x\in\mathbb{B}_R}\sum_{g\in\mathbb{B}_{2R}} 2R\kappa^{-1}\sum_{i=1}^{|g|}(f(xg_i)-f(xg_{i-1}))^2 p(xg_{i-1},xg_i)$$
$$\leq (2R)^2\kappa^{-1}|\mathbb{B}_{2R}|\mathcal{D}_{\mathbb{B}_{3R}}(f,f).$$

□

Proof of Proposition 14.4.2. As in the proof of the Poincaré inequality, it suffices to consider the case $z = 1$. Let $\varphi = \varphi_R : \Gamma \to [0,1]$ be a function that takes the value 1 on the ball $\mathbb{B}_R$ and the value 0 on $\mathbb{B}_{2R}^c$, and is Lipschitz with Lipschitz constant $1/R$. (One such function is $\varphi_R(x) = (2R - |x|)/R$ for $x \in \mathbb{B}_{2R} \setminus \mathbb{B}_R$.) Proposition 5.4.2 implies that for any such φ and any function h

$$\mathcal{D}_{\mathbb{B}_R}(h,h) \leq \mathcal{D}(h\varphi,h\varphi) = \sum_x\sum_y p(x,y)\,\{h(x)\varphi(x)-h(y)\varphi(y)\}^2 .$$

Expanding the square, using the symmetry of the kernel $p(x,y)$ and the hypothesis that h is harmonic, we obtain

$$\begin{aligned}
&\mathcal{D}(h\varphi,h\varphi)\\
&\quad=\sum_x\sum_y p(x,y)\left\{h(x)^2\varphi(x)^2-2h(x)h(y)\varphi(x)\varphi(y)+h(y)^2\varphi(y)^2\right\}\\
&\quad=2\sum_x\sum_y p(x,y)h(x)^2\varphi(x)^2-2\sum_x\sum_y p(x,y)h(x)h(y)\varphi(x)\varphi(y)\\
&\quad=2\sum_x\sum_y p(x,y)h(x)^2\varphi(x)^2-2\sum_x\sum_y p(x,y)h(x)h(y)\varphi(x)^2\\
&\qquad+2\sum_x\sum_y p(x,y)h(x)h(y)\varphi(x)(\varphi(x)-\varphi(y))\\
&\quad=2\sum_x\sum_y p(x,y)h(x)h(y)\varphi(x)(\varphi(x)-\varphi(y)).
\end{aligned}$$

(The last equality is where the harmonicity of h is needed.) Now substitute $\varphi(x) = (\varphi(x)-\varphi(y))+\varphi(y)$ and once again use the symmetry of $p(x,y)$ to deduce

$$\begin{aligned}
\mathcal{D}(h\varphi,h\varphi)&=2\sum_x\sum_y p(x,y)h(x)h(y)(\varphi(x)-\varphi(y))^2\\
&\quad+2\sum_x\sum_y p(x,y)h(x)h(y)\varphi(y)(\varphi(x)-\varphi(y))\\
&=2\sum_x\sum_y p(x,y)h(x)h(y)(\varphi(x)-\varphi(y))^2-\mathcal{D}(h\varphi,h\varphi),
\end{aligned}$$

which simplifies to

$$\mathcal{D}(h\varphi, h\varphi) = \sum_x \sum_y p(x, y)h(x)h(y)(\varphi(x) - \varphi(y))^2.$$

The inequality (14.4.2) is easily deduced from this by an application of the Cauchy-Schwarz inequality, using the fact that φ has Lipschitz constant $1/R$. Set $g(x, y) = \sqrt{p(x, y)(\varphi(x) - \varphi(y))^2}$, and observe that $g(x, y) = 0$ unless both $x, y \in \mathbb{B}_{3R}$; hence,

$$\begin{aligned}
\mathcal{D}(h\varphi, h\varphi) &= \sum_x \sum_y (g(x, y)h(x))(g(x, y)h(y)) \\
&\leq \sum_x \sum_y g(x, y)^2 h(x)^2 \\
&= \sum_{x \in \mathbb{B}_{3R}} \sum_{y \in \mathbb{B}_{3R}} p(x, y)(\varphi(x) - \varphi(y))^2 h(x)^2 \\
&\leq (1/R)^2 \sum_{x \in \mathbb{B}_{3R}} \sum_y p(x, y)h(x)^2 \\
&= (1/R)^2 \sum_{x \in \mathbb{B}_{3R}} h(x)^2.
\end{aligned}$$

□

14.5 The Colding-Minicozzi-Kleiner Theorem

The aim of this section is to prove Theorem 14.1.6. We shall assume throughout that Γ is an infinite, finitely generated group, with symmetric generating set $\mathbb{A}$. To simplify the discussion (somewhat), we shall replace the polynomial growth hypothesis (14.1.6) by the stronger assumption

$$|\mathbb{B}_m| \leq m^d \quad \text{for all large } m; \tag{14.5.1}$$

in Exercise 14.5.15 below you will be asked to check that a slight modification of the argument will prove Theorem 14.1.6 under the original hypothesis (14.1.6). Assume further that the step distribution μ of the random walk satisfies

$$\mu(x) > 0 \quad \text{if and only if} \quad x \in \mathbb{A}.$$

Thus, $\kappa := \min_{a \in \mathbb{A}} \mu(a) > 0$. Theorem 14.1.6 can be restated in the following stronger form.

Proposition 14.5.1 *There exist constants $\delta = \delta(d) < \infty$ such that for any finitely generated group Γ satisfying* (14.5.1) *and any symmetric, nearest neighbor random walk on Γ,*

$$dim(\mathcal{H}_L) \le \delta(d). \tag{14.5.2}$$

The proof will be constructive, in the sense that careful tracking of the various constants involved would yield an explicit upper bound for $\delta(d)$. In the interest of simplicity we will not attempt to do this. Because the polynomial growth rate d of the group does not depend on the choice of the generating set $\mathbb{A}$ (cf. Exercise 14.1.5), Proposition 14.5.1 shows that the bound $\delta(d)$ on the dimension of $\mathcal{H}_L$ is the same for all symmetric random walks on Γ with finitely supported step distributions.

For each positive integer R denote by Q_R the restriction to the ball $\mathbb{B}_R = \mathbb{B}_R(1)$ of the natural inner product on $L^2(\Gamma)$, that is, for any two functions $u, v : \Gamma \to \mathbb{R}$,

$$Q_R(u, v) = \sum_{x \in \mathbb{B}_R} u(x)v(x) = \left\langle u\mathbf{1}_{\mathbb{B}_R}, v\mathbf{1}_{\mathbb{B}_R} \right\rangle. \tag{14.5.3}$$

To prove Proposition 14.5.1, we will show that there exists $\delta < \infty$ such that for every set $\{u_i\}_{i \le D} \subset \mathcal{H}_L$ of centered, non-constant Lipschitz harmonic functions with cardinality $D > \delta$ and every $R \in \mathbb{N}$,

$$\det(\Psi_R) = 0 \quad \text{where} \quad \Psi_R := (Q_R(u_i, u_j))_{i,j \le D}. \tag{14.5.4}$$

This will imply that the restrictions of the functions u_i to any ball $\mathbb{B}_R$ of finite radius are linearly dependent; the assertion (14.5.2) will then follow routinely.

The remainder of this section will be devoted to the proof of (14.5.4). Throughout the discussion, we shall assume that $u_1, u_2, \cdots, u_D$ are non-constant harmonic functions on Γ satisfying $u_i(1) = 0$, all with Lipschitz constant 1, that is,

$$\sup_{x \in \Gamma} \max_{a \in A} |u_i(x) - u_i(xa)| = 1.$$

The strategy will be to show that if D is large then $\det(\Psi_R) \ll \det(\Psi_{16R})$, at least for most large radii R. Iteration of this inequality, together with the following crude *a priori* bound on the growth rate of $\det(\Psi_{16^n R})$, will then show that $\det(\Psi_R)$ must be 0.

Lemma 14.5.2 *For all R large enough that* (14.5.1) *holds,*

$$\det(\Psi_R) \le D!(R^2|\mathbb{B}_R|)^D \le D!R^{dD+2D} \tag{14.5.5}$$

Proof. Since each function u_i has Lipschitz constant 1 and satisfies $u_i(1) = 0$, it cannot exceed R in absolute value on $\mathbb{B}_R$, and hence its squared L^2-norm $Q_R(u_i, u_i)$ cannot exceed $R^2|\mathbb{B}_R|$. Consequently, by Cauchy-Schwarz, the entries

$Q_R(u_i, u_j)$ of the matrix Ψ_R cannot exceed $R^2|\mathbb{B}_R|$ in absolute value. Therefore, the determinant cannot exceed $D!(R^2|\mathbb{B}_R|)^D$. □

Bounding the ratio $\det(\Psi_R)/\det(\Psi_{16R})$ will be accomplished with the aid of the Poincaré and Cacciopoli inequalities. Use of the Poincaré inequality will require a brief digression into the study of efficient coverings of large balls $\mathbb{B}_R$ by smaller ones $\mathbb{B}_{\varepsilon R}(x_i)$, and the problem of relating the determinants $\det(\Psi_R)$ to the quadratic forms Q_R will require some elementary information from the theory of symmetric, positive semi-definite matrices.

14.5.1 A. Preliminaries: Positive Semi-Definite Matrices

A (real) $D \times D$ matrix Φ is said to be *positive semi-definite* if it is symmetric and satisfies $u^T \Phi u \geq 0$ for all $u \in \mathbb{R}^D$. (Here points $u \in \mathbb{R}^D$ are viewed as *column vectors*; for any such u we denote by u^T the *transpose* of u.) If the inequality $u^T \Phi u \geq 0$ is strict for all nonzero vectors u then Φ is *positive definite*. If Φ is positive definite, then the associated quadratic form $\langle u, u \rangle_\Phi := u^T \Phi u$ is an *inner product* on $\mathbb{R}^D$. The spectral theorem for self-adjoint linear transformations (applied with the standard inner product $\langle u, u \rangle := \langle u, u \rangle_I$) implies that for any symmetric matrix Φ there is an orthonormal basis of eigenvectors; if Φ is also positive semi-definite then the eigenvalues are nonnegative (cf. Herstein [65], Section 6.10 or Horn & Johnson [67], Section 2.5). Moreover, the determinant of a symmetric matrix is the product of its eigenvalues. There is a natural partial order, called the *Loewner order* (cf. Horn & Johnson [67], Section 7.7), on the set of nonnegative definite $D \times D$ matrices: for any two such matrices Φ, Ψ, we say that Φ dominates Ψ in the Loewner order if $\Phi - \Psi$ is also positive semi-definite, that is, if $u^T \Phi u \geq u^T \Psi u$ for all $u \in \mathbb{R}^D$.

Each of the matrices Ψ_R defined by (14.5.4) is positive semi-definite, and, since the functions u_i are linearly independent, Ψ_R is positive definite for all sufficiently large R. Furthermore, if $R_1 \leq R_2$ then the matrix Ψ_{R_2} dominates Ψ_{R_1} in the natural ordering of positive definite matrices.

Lemma 14.5.3 *If Φ, Ψ are symmetric, positive semi-definite $D \times D$ matrices such that Φ dominates Ψ in the Loewner order then* $\det(\Psi) \leq \det(\Phi)$.

Exercise 14.5.4 Prove this.

HINT: First use Gram-Schmidt to prove *Hadamard's theorem*: for any positive semi-definite matrix Φ and any orthonormal basis $e_1, e_2, \cdots, e_D$,

$$\det \Phi \leq \prod_{i=1}^{D} \|\Phi e_i\|, \tag{14.5.6}$$

with equality holding if and only if the vectors Φe_i are mutually orthogonal. Then prove Lemma 14.5.3 by using Hadamard's theorem with a suitable orthonormal basis. Alternatively, see Horn & Johnson [67], Section 7.8.

Corollary 14.5.5 *If* $R_1 \leq R_2$ *then* $\det(\Psi_{R_1}) \leq \det(\Psi_{R_2})$. □

Lemma 14.5.6 *Let* Φ, Ψ *be two positive definite* $D \times D$ *matrices such that* Φ *dominates* Ψ *in the Loewner order. Suppose that there is a vector subspace* V *of* $\mathbb{R}^D$ *such that for some* $\varepsilon < 1$

$$v^T \Psi v \leq \varepsilon^2 v^T \Phi v \quad \textit{for all } v \in V. \tag{14.5.7}$$

Then

$$\det(\Psi) \leq \varepsilon^{2k} \det(\Phi) \quad \textit{where} \quad k = \textit{dim}(V). \tag{14.5.8}$$

Proof. The Spectral Theorem implies that any positive definite matrix Σ has a positive definite square root $\Sigma^{1/2}$, that is, a positive definite matrix $\Sigma^{1/2}$ whose square is Σ. To see this, let $(u_i)_{i \leq D}$ be an orthonormal basis of eigenvectors of Σ, with corresponding eigenvalues λ_i. Since Σ is positive definite, each eigenvalue λ_i is positive, and hence has a positive square root. Set

$$\Sigma^{1/2} = \sum_{i=1}^{D} \sqrt{\lambda_i} u_i u_i^T;$$

this is clearly positive definite, and its square is Σ. Since $\Sigma^{1/2}$ is positive definite, it is invertible, with (positive definite) inverse $\Sigma^{-1/2} = \sum \lambda_i^{-1/2} u_i u_i^T$.

Let $(v_i)_{i \leq D}$ be an orthonormal basis for the inner product $\langle \cdot, \cdot \rangle_\Phi$ whose first k entries $(v_i)_{i \leq k}$ span the subspace V. (Existence of such an orthonormal basis follows by the Gram-Schmidt algorithm, using the inner product $\langle \cdot, \cdot \rangle_\Phi$.) For each index i define $w_i = \Phi^{1/2} v_i$; then for any two indices $i, j \leq D$,

$$\begin{aligned} \langle w_i, w_j \rangle &= \left\langle \Phi^{1/2} v_i, \Phi^{1/2} v_j \right\rangle \\ &= v_i^T \Phi^{1/2} \Phi^{1/2} v_j \\ &= \langle v_i, v_j \rangle_\Phi, \end{aligned}$$

so $(w_i)_{i \leq D}$ is an orthonormal basis relative to the standard inner product. Consequently, by Hadamard's inequality (14.5.6) and the multiplicativity of the determinant,

$$\frac{\det(\Psi)^{1/2}}{\det(\Phi)^{1/2}} = |\det(\Psi^{1/2} \Phi^{-1/2})|$$

$$
\begin{aligned}
&\le \prod_{i=1}^{D} \left\| \Psi^{1/2} \Phi^{-1/2} w_i \right\| \\
&= \prod_{i=1}^{D} \left\| \Psi^{1/2} v_i \right\| \\
&= \prod_{i=1}^{D} \langle v_i, v_i \rangle_{\Psi}^{1/2} \\
&\le \varepsilon^k \prod_{i=1}^{D} \langle v_i, v_i \rangle_{\Phi}^{1/2} = \varepsilon^k
\end{aligned}
$$

Here we have used the hypothesis (14.5.7), which implies that $\langle v_i, v_i \rangle_{\Psi} \le \varepsilon^2 \langle v_i, v_i \rangle_{\Phi}$ for each $i \le k$. The inequality $\langle v_i, v_i \rangle_{\Psi} \le \langle v_i, v_i \rangle_{\Phi}$, for $k+1 \le i \le D$, follows from the hypothesis that Φ dominates Ψ in the Loewner order. □

14.5.2 B. Preliminaries: Efficient Coverings

Definition 14.5.7 A subset F of a metric space $(\mathcal{X}, d)$ is said to be ρ-*separated* if no two distinct elements of F are at distance $\le \rho$. A ρ-separated subset F of $(\mathcal{X}, d)$ is *maximal* if every point $x \in \mathcal{X} \setminus F$ is at distance $\le \rho$ of F.

Obviously, if F is a maximal ρ-separated set then the collection of (closed) ρ-balls $\left\{\mathbb{B}_\rho(x)\right\}_{x \in F}$ covers the metric space $\mathcal{X}$.

Definition 14.5.8 The *intersection multiplicity* $\iota(\mathcal{U})$ of a collection of sets $\mathcal{U} = \{U_i\}_{i \in I}$ is the maximum integer m such that $\cap_{i \in J} U_i \neq \emptyset$ for some subset $J \subset I$ of cardinality m. (If no such integer m exists then the intersection multiplicity is defined to be $+\infty$.)

We now restrict attention to coverings of balls in a finitely generated group Γ, equipped with the word metric. Because the word metric is invariant (cf. equation (1.2.6)), any ball $\mathbb{B}_R(x)$ of radius R is a translate of the ball $\mathbb{B}_R$ centered at the group identity; consequently, any two such balls have the same cardinality.

Lemma 14.5.9 *Let $R_1 \le R_2$ and let $F = \{x_i\}_{i \le m}$ be a maximal R_1-separated subset of $\mathbb{B}_{R_2}$. Set $\mathcal{U}^* = \left\{\mathbb{B}_{9R_1}(x_i)\right\}_{i \le m}$. Then*

$$
|F| \le \frac{|\mathbb{B}_{2R_2}|}{|\mathbb{B}_{R_1/2}|} \quad \text{and} \tag{14.5.9}
$$

$$
\iota(\mathcal{U}^*) \le \frac{|\mathbb{B}_{32R_1}|}{|\mathbb{B}_{R_1/2}|}. \tag{14.5.10}
$$

Proof. Distinct points $x_i, x_j \in F$ are at distance greater than R_1, so no two of the balls $\mathbb{B}_{R_1/2}(x_i)$ overlap, and since $R_1 \leq R_2$, all of them are contained in $\mathbb{B}_{2R_2}$. Thus,

$$\sum_{x_i \in F} |\mathbb{B}_{R_1/2}(x_i)| \leq |\mathbb{B}_{2R_2}|.$$

Since each ball $\mathbb{B}_{R_1/2}(x_i)$ has cardinality $|\mathbb{B}_{R_1/2}|$, the inequality (14.5.9) follows.

Next, let $G \subset F$ be a subset such that $\cap_{y \in G} \mathbb{B}_{9R_1}(y) \neq \emptyset$. Let z be a point in this intersection; then each $y \in G$ is within distance $9R_1$ of z. Consequently, any two points $y, y' \in G$ are at distance $\leq 18R_1$, and therefore $G \subset \mathbb{B}_{18R_1}(y_*)$ for any $y_* \in G$. But for any two distinct points $y, y' \in G$ the balls $\mathbb{B}_{R_1/2}(y)$ and $\mathbb{B}_{R_1/2}(y')$ do not overlap, since y, y' are elements of a maximal R_1-separated set. Therefore, since each of these balls is contained in $\mathbb{B}_{19R_1}(y_*) \subset \mathbb{B}_{32R_1}$,

$$\sum_{y \in G} |\mathbb{B}_{R_1/2}(y)| \leq |\mathbb{B}_{32R_1}| \quad \Longrightarrow \quad |G| \leq \frac{|\mathbb{B}_{32R_1}|}{|\mathbb{B}_{R_1/2}|}.$$

This proves (14.5.10). □

Note: The proof shows that the bound $|\mathbb{B}_{32R_1}|/|\mathbb{B}_{R_1/2}|$ could be improved to $|\mathbb{B}_{19R_1}|/|\mathbb{B}_{R_1/2}|$, but powers of 2 are more convenient for iteration.

14.5.3 C. The Key Estimate

The *a priori* estimate of Lemma 14.5.2 shows that $\det(\Psi_R)$ grows no faster than polynomially in R. To prove Proposition 14.5.1, we will show that if D is sufficiently large then the ratio $\det(\Psi_R)/\det(\Psi_{16R})$ is bounded by a small constant not depending on R (at least for "most" R); iteration, together with the polynomial growth bound, will then show that $\det(\Psi_R)$ must be 0.

Lemma 14.5.10 *Fix $R \geq 1$ and $\varepsilon \in (0, \frac{1}{9})$, and let F be a maximal (εR)-separated subset of $\mathbb{B}_{2R}$. If $f : \Gamma \to \mathbb{R}$ is harmonic and has mean 0 on each ball $\mathbb{B}_{\varepsilon R}(x)$, where $x \in F$, then*

$$\mathcal{Q}_R(f, f) \leq \varepsilon^2 \kappa^{-1} \left(\frac{|\mathbb{B}_{32\varepsilon R}|}{|\mathbb{B}_{\varepsilon R/2}|} \right)^2 \mathcal{Q}_{16R}(f, f). \tag{14.5.11}$$

Proof. Since F is a maximal εR-separated subset of $\mathbb{B}_{2R}$, the collection $\{\mathbb{B}_{\varepsilon R}(x)\}_{x \in F}$ covers $\mathbb{B}_{2R}$, and hence, by the Poincaré inequality (14.4.1),

$$\mathcal{Q}_R(f, f) \leq \sum_{x \in F} \left\| f \mathbf{1}_{\mathbb{B}_{\varepsilon R}(x)} \right\|_2^2 \leq 2(\varepsilon^2 R^2) \kappa^{-1} \frac{|\mathbb{B}_{2\varepsilon R}|}{|\mathbb{B}_{\varepsilon R}|} \sum_{x \in F} \mathcal{D}_{\mathbb{B}_{3\varepsilon R}(x)}(f, f).$$

(The second inequality depends on the hypothesis that f has mean 0 on each of the balls $\mathbb{B}_{\varepsilon R}(x)$.) Lemma 14.5.9 implies that the intersection multiplicity of the covering $\{\mathbb{B}_{3\varepsilon R}(x)\}_{x\in F}$ does not exceed $|\mathbb{B}_{32\varepsilon R}|/|\mathbb{B}_{\varepsilon R/2}|$, so no edge of the Cayley graph lies in more than $|\mathbb{B}_{32\varepsilon R}|/|\mathbb{B}_{\varepsilon R/2}|$ of the balls $\mathbb{B}_{3\varepsilon R}(x)$, where $x \in F$. Since the balls $\mathbb{B}_{3\varepsilon R}(x)$ are all contained in $\mathbb{B}_{3R}$ (by the hypothesis that $\varepsilon < 1/9$), it follows that

$$\sum_{x\in F} \mathcal{D}_{\mathbb{B}_{3\varepsilon R}(x)}(f, f) \leq \frac{|\mathbb{B}_{32\varepsilon R}|}{|\mathbb{B}_{\varepsilon R/2}|}\mathcal{D}_{\mathbb{B}_{3R}}(f, f).$$

The Caccioppoli inequality (14.4.2) therefore implies that

$$\begin{aligned}
Q_R(f, f) &\leq 2\varepsilon^2 R^2 \kappa^{-1} \left(\frac{|\mathbb{B}_{\varepsilon R}|}{|\mathbb{B}_{\varepsilon R/2}}\right)\left(\frac{|\mathbb{B}_{32\varepsilon R}|}{|\mathbb{B}_{\varepsilon R/2}|}\right)(3R)^{-2} Q_{9R}(f, f) \\
&\leq \frac{2}{9}\varepsilon^2 \kappa^{-1}\left(\frac{|\mathbb{B}_{32\varepsilon R}|}{|\mathbb{B}_{\varepsilon R/2}|}\right)^2 Q_{9R}(f, f) \\
&\leq \varepsilon^2 \kappa^{-1}\left(\frac{|\mathbb{B}_{32\varepsilon R}|}{|\mathbb{B}_{\varepsilon R/2}|}\right)^2 Q_{16R}(f, f).
\end{aligned}$$

□

Corollary 14.5.11 *If F is a maximal (εR)-separated subset of $\mathbb{B}_{2R}$, where $\varepsilon < 1/9$, then*

$$\det(\Psi_R) \leq \left(\varepsilon^2 \kappa^{-1}\left(\frac{|\mathbb{B}_{32\varepsilon R}|}{|\mathbb{B}_{\varepsilon R/2}|}\right)^2\right)^{D-|F|} \det(\Psi_{16R}) \tag{14.5.12}$$

Proof. Let V be the vector subspace of $\mathrm{span}(u_i)_{i\leq D}$ consisting of those functions f in the linear span that have mean 0 on each ball $\mathbb{B}_{\varepsilon R}(x_i)$, where $x_i \in F$. This subspace is determined by $|F|$ linear constraints, so its dimension is at least $D-|F|$. Therefore, by Lemma 14.5.6, the inequality (14.5.12) follows from (14.5.11). □

Remark 14.5.12 Inequality (14.5.12) holds for every integer $D \geq 1$. This will be of critical importance in the proof of Proposition 14.5.1.

14.5.4 D. Bounded Doubling

Definition 14.5.13 A finitely generated group Γ (or its Cayley graph) is said to have the *bounded doubling property* if there exists $C < \infty$ such that for every real number $R > 0$,

$$\frac{|\mathbb{B}_{2R}|}{|\mathbb{B}_R|} \leq C. \tag{14.5.13}$$

The bounded doubling property implies polynomial growth, with polynomial growth exponent $d = \log_2 C$. A theorem of Bass [7] and Guivarc'h [62] implies that the converse is also true. The proof of this theorem, however, relies on Gromov's classification of such groups, a result we will prove in Chapter 15 with the aid of the Colding-Minicozzi-Kleiner theorem. Therefore, we cannot appeal to the Bass-Guivarc'h theorem in proving Proposition 14.5.1. Nevertheless, it is instructive to see how the Key Estimate (Corollary 14.5.11) works in the presence of bounded doubling.

Proof of Proposition 14.5.1 for Groups with Bounded Doubling. Assume that the group Γ satisfies inequality (14.5.13), with $C = 2^K$ for some positive integer K. By Lemma 14.5.9, for any $R > 0$ and any $\varepsilon < 1/9$ the cardinality of any (εR)-separated subset F of $\mathbb{B}_{2R}$ satisfies

$$|F| \leq \frac{|\mathbb{B}_{4R}|}{|\mathbb{B}_{\varepsilon R/2}|} \leq C^{2+\log_2(8/\varepsilon)} = (32)^K \varepsilon^{-K}. \tag{14.5.14}$$

Moreover, by Corollary 14.5.11, for any $R > 0$ and $\varepsilon < 1/9$,

$$\frac{\det(\Psi_R)}{\det(\Psi_{16R})} \leq \left(\varepsilon^2 \kappa^{-1} \left(\frac{|\mathbb{B}_{32\varepsilon R}|}{|\mathbb{B}_{\varepsilon R/2}|}\right)^2\right)^{D-|F|} \leq (2^{12K}\kappa^{-1}\varepsilon^2)^{D-|F|}. \tag{14.5.15}$$

By iteration, we conclude that with $\tilde{\varepsilon} := 2^{12K}\kappa^{-1}\varepsilon^2$ and $D_* = (32)^K\varepsilon^{-K}$, for every integer $n \geq 1$,

$$\frac{\det(\Psi_R)}{\det(\Psi_{16^n R})} \leq \tilde{\varepsilon}^{nD-nD_*}, \tag{14.5.16}$$

and so by Lemma 14.5.2,

$$\det(\Psi_R) \leq D!(16^n R)^{(d+2)D}\tilde{\varepsilon}^{nD-nD_*}. \tag{14.5.17}$$

Inequality (14.5.17) is valid for every $\varepsilon \in (0, \frac{1}{9})$ and all positive integers n, D. Thus, by fixing $\varepsilon > 0$ sufficiently small that $16^{(d+2)}\tilde{\varepsilon} < 1$, we can arrange that the bound (14.5.17) decays exponentially with n by taking D sufficiently large. In particular, for any $\varepsilon > 0$ small enough that $16^{(d+2)}\tilde{\varepsilon} < 1$, there exists an integer $D > D_* = D_*(\varepsilon)$ such that

$$16^{(d+2)D}\tilde{\varepsilon}^{D-D_*} < 1.$$

This, together with (14.5.17), implies that $\det(\Psi_R) = 0$. Consequently, no linearly independent set of functions in $\mathcal{H}_L$ can have cardinality as large as D. □

14.5.5 E. The General Case

The essential idea behind the proof of Proposition 14.5.1 in the general case is the same as in the special case considered above, but the argument is complicated by the fact that the doubling inequality (14.5.13) need not hold at all scales. However, if the group Γ has polynomial growth with exponent d then the inequality (14.5.13) cannot fail at too many different scales R for $C > 2^d$. The following lemma provides a quantification of this.

Lemma 14.5.14 *Assume that the group* Γ *satisfies the polynomial growth hypothesis* (14.5.1) *with exponent* d*. Fix integers* $J, K, L, R \geq 1$*, and for each positive integer* M *define*

$$N_M = \left| \left\{ 0 \leq m < M \, : \, \frac{|\mathbb{B}_{2^{mK+LK} \times R}|}{|\mathbb{B}_{2^{mK} \times R}|} > 2^{JKLd} \right\} \right| . \tag{14.5.18}$$

Then there exists a constant $C = C(J, K, L, R, d) < \infty$ *such that for all sufficiently large* $M \in \mathbb{N}$*,*

$$N_M \leq \frac{ML}{J} + C. \tag{14.5.19}$$

Proof. Consider first the special case $L = 1$. The cardinalities $\mathbb{B}_n$ are nondecreasing in n, so for every $K \in \mathbb{N}$,

$$\frac{|\mathbb{B}_{2^{mK+K} \times R}|}{|\mathbb{B}_{2^{mK} \times R}|} \geq 1.$$

Therefore,

$$|\mathbb{B}_{2^{MK} \times R}| = |\mathbb{B}_R| \prod_{m=0}^{M-1} \frac{|\mathbb{B}_{2^{mK+K} \times R}|}{|\mathbb{B}_{2^{mK} \times R}|} \geq 2^{JKdN_M} |\mathbb{B}_R| \geq 2^{JKdN_M},$$

and so the polynomial growth hypothesis implies

$$2^{JKdN_M} \leq |\mathbb{B}_{2^{MK} \times R}| \leq 2^{dMK} R^d.$$

The inequality (14.5.19) obviously follows (with $C = (\log_2 R)/(JdK)$). This proves (14.5.19) for $L = 1$. The general case, where $L \geq 1$ is any positive integer, now follows by applying the special case L times, with K replaced by KL and R replaced successively by $R, 2^K R, 2^{2K} R, \cdots, 2^{LK-K} R$. □

Proof of Proposition 14.5.1. First, because Lemma 14.5.3 ensures that the determinants $\det \Psi_R$ are nondecreasing in R, it suffices to establish that if D is large then (14.5.4) holds for all large powers $R = 2^S$ of 2. Set

$$\varepsilon = (16)^{-\nu}$$

where ν is a positive integer whose value will be fixed later, and fix $S \in \mathbb{N}$ sufficiently large that $\frac{1}{2}\varepsilon R \in \mathbb{N}$. For each $m = 0, 1, 2, \cdots$, let F_m be a maximal $(\varepsilon \times 16^m R)$-separated subset of $\mathbb{B}_{2\times(16^m R)}$. By Lemma 14.5.9,

$$|F_m| \leq \frac{|\mathbb{B}_{4\times 16^m R}|}{|\mathbb{B}_{\varepsilon\times 16^m R/2}|} \leq \frac{|\mathbb{B}_{8\times 16^m R}|}{|\mathbb{B}_{\varepsilon\times 16^m R/2}|} \tag{14.5.20}$$

(the second bound has the advantage that the ratio of the two radii is a power of 16), and by Corollaries 14.5.5 and 14.5.11,

$$\frac{\det(\Psi_{16^m R})}{\det(\Psi_{16^{m+1} R})} \leq \left\{ 1 \wedge \left(\varepsilon^2 \kappa^{-1} \left(\frac{|\mathbb{B}_{(16)^2 \varepsilon\times 16^m R/2}|}{|\mathbb{B}_{\varepsilon\times 16^m R/2}|} \right)^2 \right)^{D-|F_m|} \right\}. \tag{14.5.21}$$

(Here we have used the fact that $|\mathbb{B}_{32\varepsilon\times(16)^m R}| \leq |\mathbb{B}_{\varepsilon\times 16^{m+2} R/2}|$; once again, the reason for using the cruder bound is that the ratio of the radii is a power of 16.)

Fix $J \in \mathbb{N}$, and for each $M = 1, 2, 3, \cdots$, define

$$N'_M = \left| \left\{ 1 \leq m \leq M \,:\, \frac{|\mathbb{B}_{(16)^2 \varepsilon\times 16^m R/2}|}{|\mathbb{B}_{\varepsilon\times 16^m R/2}|} > (16)^{2Jd} \right\} \right| \quad \text{and}$$
$$N''_M = \left| \left\{ 1 \leq m \leq M \,:\, \frac{|\mathbb{B}_{8\times 16^m R}|}{|\mathbb{B}_{\varepsilon\times 16^m R/2}|} > (16/\varepsilon)^{Jd} \right\} \right|.$$

By Lemma 14.5.14 (with $K = 4$), there exists $J \in \mathbb{N}$ not depending on ε or R such that for all sufficiently large M,

$$N'_M + N''_M < \frac{M}{4} + \frac{M}{4} = \frac{M}{2}.$$

Consequently, by inequalities (14.5.20) and (14.5.21), for all large M,

$$\frac{\det(\Psi_R)}{\det(\Psi_{16^M R})} = \prod_{m=0}^{M-1} \frac{\det(\Psi_{16^m R})}{\det(\Psi_{16^{m+1} R})} \leq \left(\varepsilon^2 \kappa^{-1} \times (16)^{2Jd} \right)^{(D-D')\times M/2} \tag{14.5.22}$$

where

$$D' = (16/\varepsilon)^{Jd}.$$

Inequality (14.5.22), together with the crude bound (14.5.5) on $\det(\Psi_{16^{m+1} R})$, implies that for all sufficiently large M,

$$\det(\Psi_R) \leq (\varepsilon^2 \kappa^{-1} \times (16)^{2Jd})^{(D-D')\times M/2} \times (D!(16^M R)^{dD+2D})$$

$$\leq C_{R,D} \times \varepsilon^{(D-D')M} \times (16)^{Jd(D-D')M+(dD+2D)M} \tag{14.5.23}$$

where $C_{R,D}$ is a constant depending on R, D, κ, and the polynomial growth rate d but not on M. Hence, if $D > 2D'$ then for all large M,

$$\det(\Psi_R) \leq C_{R,D} \times \varepsilon^{MD/2} \times (16)^{JdDM/2+(dD+2D)M},$$

and so if $\varepsilon = (16)^{-\nu}$ with exponent ν large enough that

$$\nu D/2 > JdD/2 + (d+2)D$$

then for all $D > 2D'$ we must have $\det(\Psi_R) = 0$. Thus, no linearly independent set of non-constant harmonic functions that all vanish at the identity 1 can have cardinality larger than $2D' = 2 \times (16)^{(1+\nu)Jd}$. □

Exercise 14.5.15 Show how the preceding proof can be adapted to prove the Colding-Minicozzi-Kleiner theorem under the weaker hypothesis 14.1.6.

HINT: The conclusion (14.5.19) of Lemma 14.5.14 will hold for any $M \in \mathbb{N}$ such that $|\mathbb{B}_{2^{MK}}| \leq 2^{dMK}$. Show that if (14.1.6) holds then (for a slightly larger d) there will be an infinite sequence of such "good" M.

Exercise 14.5.16 Prove that under the hypotheses of Theorem 14.1.6, for each integer m the vector space $\mathcal{H}^m$ of centered harmonic functions u satisfying

$$|u(x)| \leq C_u |x|^m$$

is finite-dimensional.

HINT: The only place in the foregoing argument where the Lipschitz condition was used was Lemma 14.5.2.

Additional Notes. The main result of this chapter, Theorem 14.1.6, is due to Kleiner [80]. It is the discrete analogue of an important result of Colding and Minicozzi [26] describing spaces of harmonic functions satisfying various growth conditions. For any integer $m \geq 1$, define $\mathcal{H}_L^m$ to be the vector space of harmonic functions with *polynomial growth* of exponent $\leq m$, that is, harmonic functions h such that for some constant $C_h < \infty$,

$$|h(x) - h(1)| \leq C_h |x|^m \quad \text{for all} \quad x \in \Gamma. \tag{14.5.24}$$

(Note: The absolute value $|h(x)|$ refers to the usual absolute value function on $\mathbb{R}$, whereas $|x| = |x|_{\mathbb{A}}$ refers to the word norm on Γ.) Then for each $m \in \mathbb{N}$ the vector space $\mathcal{H}_L^m$ is finite-dimensional, with dimension depending on m.

The study of polynomial-growth harmonic functions on complete Riemannian manifolds has been an important theme of global analysis since the 1980s, when

S.-T. Yau first conjectured that, for an open Riemannian manifold with nonnegative Ricci curvature, the space of harmonic functions with prescribed polynomial growth rate is finite-dimensional. Yau's conjecture was ultimately proved by Colding and Minicozzi [26], following earlier partial results by Li and Tam [89]. Subsequently, Kleiner [80], using similar techniques, proved the corresponding result for finitely generated groups; Theorem 14.1.6 is a special case. Various among these authors also obtained bounds on the dimensions of these spaces depending only on the dimensionality of the underlying manifold; Li and Tam, for instance, showed that for a complete manifold with nonnegative Ricci curvature and volume growth $|\mathbb{B}_R| \leq CR^n$ the space of linear-growth harmonic functions has dimension at most $n + 1$. For a thorough account of these developments, see Li's book [88].

The exposition in Sections 14.3 and 14.5 follows the strategy sketched by T. Tao in his blog article [122].

Chapter 15
Groups of Polynomial Growth

15.1 The Kleiner Representation

In this chapter we will explore some of the structural consequences of the Colding-Minicozzi-Kleiner theorem (Theorem 14.1.6) for finitely generated groups of polynomial growth. This theorem asserts that for any such group Γ, the space $\mathcal{H}_L$ of centered Lipschitz harmonic functions (for any symmetric step distribution charging every element of a symmetric generating set) is finite-dimensional. Thus, the left translation operators define a *finite-dimensional* linear representation of Γ. The existence of this representation has far-reaching consequences, as we will see.

Assumption 15.1.1 *Assume in this section that Γ is an infinite, finitely generated group of polynomial growth with finite, symmetric generating set $\mathbb{A}$. Let μ be a symmetric probability distribution with exact support $\mathbb{A}$ (that is, $\mu(a) > 0$ if and only if $a \in \mathbb{A}$), and let $D = dim(\mathcal{H}_L)$, where $\mathcal{H}_L$ is the real vector space of centered Lipschitz μ-harmonic functions.*

Recall that Γ has polynomial growth if there exist constants $C, d < \infty$ such that for infinitely many integers $m \geq 1$ the volume $|\mathbb{B}_m|$ of the ball $\mathbb{B}_m$ of radius m centered at the group identity does not exceed Cm^d. The *degree* of polynomial growth is defined to be

$$d := \liminf_{m\to\infty} \frac{\log |\mathbb{B}_m|}{\log m}. \tag{15.1.1}$$

Clearly, there are no infinite, finitely generated groups with polynomial growth of degree < 1, because for an infinite group the ball $\mathbb{B}_m$ must always have cardinality at least m.

For any centered ($\mu-$) harmonic function $h : \Gamma \to \mathbb{R}$ and any group element x, the *left translate* and *centered left translate* are defined by

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5_15

$$\mathcal{L}_x h(y) = h(x^{-1}y) \quad \text{and} \tag{15.1.2}$$

$$\mathcal{L}_x^0 h(y) = h(x^{-1}y) - h(x^{-1}). \tag{15.1.3}$$

The translation operator $\mathcal{L}_x^0$ defines an invertible, norm-preserving linear mapping $\mathcal{L}_x^0 : \mathcal{H}_L \to \mathcal{H}_L$ on the space $\mathcal{H}_L$ of centered Lipschitz harmonic functions, and by Exercise 14.1.3 (cf. in particular equations (14.1.4)) the mapping $x \mapsto \mathcal{L}_x^0$ is a homomorphism $\mathcal{L}^0 : \Gamma \to GL(\mathcal{H}_L)$ from Γ to the algebra of norm-preserving linear transformations of $\mathcal{H}_L$, that is, a *linear representation* of Γ in $GL(\mathcal{H}_L)$. Let's call this the *Kleiner representation*.

Lemma 15.1.2 *There is an inner product $\langle \cdot, \cdot \rangle$ on the vector space $\mathcal{H}_L$ that is preserved by every translation operator $\mathcal{L}_x^0$, that is, for all $u, v \in \mathcal{H}_L$ and $x \in \Gamma$,*

$$\left\langle \mathcal{L}_x^0 u, \mathcal{L}_x^0 v \right\rangle = \langle u, v \rangle . \tag{15.1.4}$$

Consequently, the mapping $x \mapsto \mathcal{L}_x^0$ is an orthogonal *representation, that is, $\mathcal{L}^0 : \Gamma \to \mathbb{O}_D$ can be viewed as a linear representation of Γ in the orthogonal group $\mathbb{O}_D$.*

Proof Sketch. This is a standard argument in the elementary theory of group representations — see, for instance, Bröcker & tom Dieck [18], Chapter II, Theorem 1.7. Since each $\mathcal{L}_x^0$ preserves the Lipschitz norm, the image $\mathcal{L}^0(\Gamma)$ in the space $GL(\mathcal{H}_L) \cong GL_m$ of $m \times m$ real matrices is bounded, and hence has compact closure K. Since the group K is compact, it has a left-invariant Borel probability measure λ, that is, a Borel probability measure ν on K that satisfies

$$\int f(h)\, d\lambda(h) = \int f(gh)\, d\lambda(h)$$

for every $g \in K$ and every continuous function $f : K \to \mathbb{R}$. (See Exercise 15.1.3 below. This measure, known as *Haar measure*, is unique, and it is also right-invariant, but these facts are irrelevant to the proof.) Take any inner product $\langle \cdot, \cdot \rangle_*$ on $\mathcal{H}_L$, and set

$$\langle u, v \rangle := \int_{k \in K} \langle ku, kv \rangle_* \, \lambda(dk);$$

then the invariance of λ implies that $\langle ku, kv \rangle = \langle u, v \rangle$ for any $k \in K$. □

Exercise 15.1.3 (Haar Measure on a Compact Group) † This exercise outlines a proof that every compact, metrizable group G has a left-invariant Borel probability measure λ. This proof is based on random walk theory, and uses an extension of the *coupling method* discussed in Section 9.6.

Let G be a compact group whose topology is induced by a metric ρ. (Thus, the group operations $(g, h) \mapsto gh$ and $g \mapsto g^{-1}$ are continuous with respect to ρ).

Let $D = \left\{g_n^{\pm 1}\right\}_{n \in \mathbb{N}} \cup \{1\}$ be a symmetric, countable, dense subset of G and let μ be a probability distribution on D that gives positive probability to every $g \in D$. Let $(X_n)_{n \geq 0}$ and $(X_n')_{n \geq 0}$ be independent random walks on G, both with step distribution μ and initial point 1.

(A) Show that there is a subsequence $(n_m)_{m \in \mathbb{N}}$ of the nonnegative integers such that for every $g \in G$ the random variables gX_n' have distributions that converge weakly along the subsequence $(n_m)_{m \in \mathbb{N}}$. Denote by λ_g the weak limits; show that these probability measures satisfy

$$\int f(gh)\, d\lambda(h) = \int f(h)\, d\lambda_g(h).$$

HINT: Helly's Selection Principle and the continuity of the group operations.

(B) Show that for any $\varepsilon > 0$ there exists $\delta > 0$ such that the measure μ gives mass at least ε to every open ball in G of radius δ.

(C) Use (B) to show that for any $g \in G$ and any $\delta > 0$ the δ-*coupling time* random variable

$$T = T(g; \delta) := \min\left\{n \in \mathbb{Z}_+ \,:\, \rho(X_n, gX_n') < \delta\right\}$$

is almost surely finite.

(D) Denote by ξ_n and ξ_n' the increments of the random walks $(X_n)_{n \geq 0}$ and $(X_n')_{n \geq 0}$, and write $T = T(g; \delta)$. Define

$$\begin{aligned} X_n'' &= gX_n' \quad \text{for all } n \leq T \quad \text{and} \\ X_n'' &= gX_T'\xi_{T+1}\xi_{T+2}\cdots\xi_n \quad \text{for all } n > T. \end{aligned}$$

Show that the sequence $(X_n'')_{n \geq 0}$ has the same law as $(gX_n')_{n \geq 0}$, and that for every $\varepsilon > 0$ there is a $\delta > 0$ sufficiently small that

$$\rho(X_n, X_n'') < \varepsilon \quad \text{for all } \; n \geq T.$$

(E) Use (D) to show that for any continuous $f : G \to \mathbb{R}$ and any subsequence $(n_m)_{m \in \mathbb{N}}$ of the nonnegative integers along which the distributions of the random variables gX_n' converge weakly, the weak subsequential limits λ, λ_g satisfy

$$\left| \int f\, d\lambda - \int f\, d\lambda_g \right| \leq \max\left\{|f(x) - f(y)| \,:\, \rho(x, y) \leq \delta\right\}.$$

(F) Conclude that for every continuous $f : G \to \mathbb{R}$ and every $g \in G$,

$$\int f\, d\lambda = \int f\, d\lambda_g.$$

The image $\mathcal{L}^0(\Gamma)$ of Γ under the Kleiner representation is a subgroup of the orthogonal group that can be either finite or infinite. In the special case where Γ is the additive group $\Gamma = \mathbb{Z}^D$, the vector space $\mathcal{H}_L$ is spanned by the coordinate functions $u_i(x) = x_i$ (cf. Exercise 14.2.1), and it is easily checked that these are invariant by translations (after centering), so the subgroup $\mathcal{L}^0(\Gamma) = \{I\}$ is trivial. This is an extreme case. Here is another; this example shows that the Kleiner representation can be trivial even for groups that are not virtually abelian.

Exercise 15.1.4 (Heisenberg Group) † The *Heisenberg group* $\mathbb{H}$ is the group of 3×3 matrices of the form

$$g = \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \quad \text{where} \quad x, y, z \in \mathbb{Z}.$$

(A) Check that $\mathbb{H}$ is generated by the four matrices

$$\begin{pmatrix} 1 & \pm 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \pm 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

(B) Check that each of the coordinate functions $u_x(g) = x$, $u_y(g) = y$, $u_z(z) = z$ is harmonic for the random walk with uniform step distribution on the 4 generators in part (A). Show that only u_x and u_z are Lipschitz.

(C) Let $V = \operatorname{span}(u_x, u_z)$ be the 2-dimensional subspace of $\mathcal{H}_L$ spanned by the coordinate functions u_x, u_z. Check that every $\mathcal{L}^0_x$, where $x \in \mathbb{H}$, acts as the identity on V.

(D) ** Show that $\mathcal{H}_L = V$, i.e., that all Lipschitz harmonic functions are linear combinations of the coordinate functions u_x, u_z.

HINT: Show that every harmonic function is a linear combination of u_x, u_y, u_z.

Exercise 15.1.5 † Let Γ be the group of 2×2 matrices of the form

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -1 & x \\ 0 & -1 \end{pmatrix} \quad \text{where} \quad x \in \mathbb{Z}.$$

(A) Show that Γ is generated by the three matrices

$$\begin{pmatrix} +1 & \pm 1 \\ 0 & +1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(B) Describe the space $\mathcal{H}_L$ of Lipschitz harmonic functions for the random walk on Γ with uniform step distribution on the generators in (A).

(C) What is the group $\mathcal{L}^0(\Gamma)$ in this case?

Proposition 15.1.6 *If the image $\mathcal{L}^0(\Gamma)$ of Γ in the orthogonal group $\mathbb{O}_D$ is finite, then the subgroup $K := ker(\mathcal{L}^0)$ has finite index in Γ, and there is a surjective homomorphism $\mathbf{u} : K \to \mathbb{Z}^D$.*

Proof. The index of K in Γ is the order of the image group $\mathcal{L}^0(\Gamma)$, so the hypothesis that $\mathcal{L}^0(\Gamma)$ is finite implies that K is a finite-index subgroup.

For every $g \in K$, the linear transformation $\mathcal{L}^0_g : \mathcal{H}_L \to \mathcal{H}_L$ is the identity. Hence, for every $u \in \mathcal{H}_L$, $g \in K$, and $x \in \Gamma$,

$$u(gx) = u(g) + u(x). \tag{15.1.5}$$

Let $\{u_i\}_{i \le D}$ be a basis for $\mathcal{H}_L$, and consider the map $\mathbf{u} : \Gamma \to \mathbb{R}^D$ defined by

$$\mathbf{u}(g) = (u_1(g), u_2(g), \cdots u_D(g)).$$

Equation (15.1.5) implies that the restriction of $\mathbf{u}$ to K is a group homomorphism. We claim that the image $\mathbf{u}(K)$ of this homorphism is an infinite additive subgroup of $\mathbb{R}^D$ of full rank (that is, it is not contained in a proper vector subspace of $\mathbb{R}^D$). To see this, suppose to the contrary that $\mathbf{u}(K)$ is contained in a proper subspace of $\mathbb{R}^D$; then some nontrivial linear combination $u := \sum_{i \le D} b_i u_i$ must be identically 0 on K. Let $\{Kx_i\}_{i \le m}$ be an enumeration of the distinct right cosets of K in Γ; then for any $y \in Kx_i$ there is an element $g \in K$ such that $y = gx_i$. Hence, since $u = 0$ on K, equation (15.1.5) implies that for $y = g_{x_i} Kx_i$,

$$u(y) = u(gx_i) = u(g) + u(x_i) = u(x_i).$$

Thus, the function u assumes only finitely distinct values on Γ. But u is harmonic, so by the Maximum Principle (cf. Proposition 6.2.7) this could only occur if u is a constant function. Since $u \in \mathcal{H}_L$, it must vanish at the group identity 1, so it follows that $u \equiv 0$. Finally, since the functions u_i are by assumption linearly independent, it follows that the coefficients b_i in the representation of u are all 0. This proves that $\mathbf{u}(K)$ is not contained in a proper subspace of $\mathbb{R}^D$. Since the additive group $\mathbb{R}^D$ has no elements of finite order (other than the group identity $\mathbf{0}$), and since $\mathbf{u}(K)$ is an additive subgroup of $\mathbb{R}^D$, it follows that the set $\mathbf{u}(K) \subset \mathbb{R}^D$ is unbounded, and therefore countably infinite.

As the subgroup K is finitely generated (cf. Exercise 1.2.11), so is $\mathbf{u}(K)$, and since $\mathbf{u}(K)$ is a subgroup of $\mathbb{R}^D$ it has no elements of finite order. Thus, $\mathbf{u}(K)$ is a finitely generated, free abelian group. By the Structure Theorem for finitely generated abelian groups (cf. Herstein [65], Theorem 4.5.1), it follows that $\mathbf{u}(K)$ is isomorphic to $\mathbb{Z}^d$ for some integer $1 \le d \le D$. We have shown that $\mathbf{u}(K)$ is not contained in any vector subspace of $\mathbb{R}^D$ of dimension $< D$; therefore $\mathbf{u}(K)$ cannot be isomorphic to $\mathbb{Z}^d$ for any $d < D$, and so

$$\mathbf{u}(K) \cong \mathbb{Z}^D. \tag{15.1.6}$$

This proves that $\mathbf{u} : K \to \mathbb{Z}^D$ is a surjective homomorphism. □

Corollary 15.1.7 *If dim*$(\mathcal{H}_L) \geq 3$ *and if* $\mathcal{L}^0(\Gamma)$ *is finite, then* Γ *is a transient group.* □

15.2 Finitely Generated Subgroups of $\mathbb{U}_D$

In general, the image $\mathcal{L}^0(\Gamma)$ of the Kleiner representation need not be finite, but it is always a finitely generated subgroup of the orthogonal group $\mathbb{O}_D$. The case $D = 2$ is particularly simple, because in this case the orthogonal group $\mathbb{O}_2$ is just the circle group $\mathbb{R}/\mathbb{Z}$, which is abelian. The main result of this section, Theorem 15.2.1, will show that even for higher dimensions $D \geq 3$, where the orthogonal group $\mathbb{O}_D$ is non-abelian, the image $\mathcal{L}^0(\Gamma)$ of Γ under the Kleiner representation, although not necessarily abelian, is always *virtually* abelian (i.e., has an abelian subgroup of finite index).

The orthogonal group $\mathbb{O}_D$ is a subgroup of the *unitary* group $\mathbb{U}_D$, which consists of all complex, $D \times D$ matrices g that satisfy $gg^* = g^*g = I$, where g^* denotes the conjugate transpose of g. Equivalently, a $D \times D$ matrix g with complex entries is unitary if it preserves the standard inner product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{D} v_i \bar{w}_i \tag{15.2.1}$$

on $\mathbb{C}^D$ (where $\bar{\alpha}$ denotes the complex conjugate of α), that is,

$$\langle g\mathbf{v}, g\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle \quad \text{for all } \mathbf{v}, \mathbf{w} \in \mathbb{C}^D. \tag{15.2.2}$$

Theorem 15.2.1 *Any finitely generated subgroup of the unitary group* $\mathbb{U}_D$ *that has polynomial growth is virtually abelian.*

Before we proceed with the proof, let's take note of the following consequence.

Corollary 15.2.2 *Let* Γ *be a finitely generated group of polynomial growth. If the image* $\mathcal{L}^0(\Gamma)$ *of* Γ *under the Kleiner representation is infinite, then* Γ *has a finite-index subgroup* Γ_0 *for which there is a surjective homomorphism* $\varphi : \Gamma_0 \to \mathbb{Z}^k$ *onto some integer lattice* $\mathbb{Z}^k$ *of dimension* $k \geq 1$.

Proof. Theorem 15.2.1 and the Structure Theorem for finitely generated abelian groups imply that $\mathcal{L}^0(\Gamma)$ has a finite-index subgroup H isomorphic to $\mathbb{Z}^k$, for some integer $k \geq 1$. Consequently, $(\mathcal{L}^0)^{-1}(H) := \Gamma_0$ is a finite-index subgroup of Γ that $\mathcal{L}^0$ maps surjectively onto H. □

The remainder of this section will be devoted to the proof of Theorem 15.2.1. The special case $D = 1$ is, of course, trivial, as the group $\mathbb{U}_1$ is abelian, so we

shall assume henceforth that $D > 1$. Two properties of unitary matrices will come into play: first, that the usual matrix norm induces an *invariant metric* on $\mathbb{U}_D$, and second, the *Spectral Theorem* for unitary matrices.

Exercise 15.2.3 Let $\|\cdot\|$ be the complex matrix norm

$$\|g\| := \max_{\mathbf{u}\in\mathbf{C}^D : \|\mathbf{u}\|=1} \|g\mathbf{u}\|, \tag{15.2.3}$$

where $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle}$, and for any two matrices $x, y \in \mathbb{U}_D$ define

$$\rho(x, y) := \|x - y\|. \tag{15.2.4}$$

(A) Verify that ρ is an invariant metric on $\mathbb{U}_D$, that is, a metric such that

$$\rho(x, y) = \rho(gx, gy) \quad \text{for all } \; g, x, y \in \mathbb{U}_D.$$

(B) Show that the groups $\mathbb{U}_D$ and $\mathbb{O}_D$ are compact in the topology induced by the metric ρ.

Spectral Theorem for Unitary Matrices. *For any $D \times D$ unitary matrix g there exist an orthonormal basis $\{u_i\}_{i\leq D}$ of $\mathbb{C}^D$ (the eigenvectors) and a corresponding set $\{\lambda_i\}_{i\leq D}$ of complex numbers, each of modulus 1 (the eigenvalues), such that for every $v \in \mathbb{C}^D$,*

$$gv = \sum_{i=1}^{D} \langle v, u_i\rangle \lambda_i u_i. \tag{15.2.5}$$

See Herstein [65], Section 6.10 or Horn & Johnson [67], Section 2.5 for the proof. The Spectral Theorem has the following elementary but important consequence.

Corollary 15.2.4 *Two $D \times D$ unitary matrices g, h commute if and only if they are simultaneously diagonalizable, that is, if and only if g and h have a common orthonormal basis of eigenvectors.*

Definition 15.2.5 The *infinite rooted binary tree* $\mathbb{T}$ is the infinite directed graph with vertex set equal to the set $\mathcal{V} = \cup_{n=1}^{\infty} \{0, 1\}^n$ of all finite binary words (including the empty word $\emptyset$, which is designated the *root*) and with directed edges

$$x_1x_2\cdots x_n \longrightarrow x_1x_2\cdots x_n0 \quad \text{and}$$
$$x_1x_2\cdots x_n \longrightarrow x_1x_2\cdots x_n1$$

from vertices to their two immediate *children*. The *depth* of a vertex $x_1x_2\cdots x_n$ is defined to be the length n of its representative word. An ε-*embedding* of $\mathbb{T}$ into a metric space $(\mathcal{X}, d)$ is a function $f : \mathcal{V} \to \mathcal{X}$, such that

(a) $f(\emptyset) \neq f(1)$;
(b) $f(x_1x_2\cdots x_n) = f(x_1x_2\cdots x_n0)$ for every vertex $x_1x_2\cdots x_n$; and
(c) for every vertex $x_1x_2\cdots x_n$,

$$0<d(f(x_1x_2\cdots x_n0), f(x_1x_2\cdots x_n01))<\varepsilon d(f(x_1x_2\cdots x_n), f(x_1x_2\cdots x_n1))$$

and

$$0<d(f(x_1x_2\cdots x_n1), f(x_1x_2\cdots x_n11))<\varepsilon d(f(x_1x_2\cdots x_n), f(x_1x_2\cdots x_n1)).$$

Lemma 15.2.6 *Let $f : \mathbb{T} \to \mathcal{X}$ be an ε-embedding of the infinite rooted binary tree for some $0 < \varepsilon < 1/3$. Then for any two vertices $v, v' \in \mathcal{V}$ such that depth(v) $\leq$depth(v')*

$$f(v) = f(v') \quad \textit{if and only if} \quad v' = v00\cdots 0,$$

that is, v' is obtained from v by adjoining a finite string of 0s. Consequently, the number of distinct points in the set $\{f(v) : \text{depth}(v) \leq n\}$ is 2^n.

Exercise 15.2.7 Prove this.

The strategy behind the proof of Theorem 15.2.1 will be to show that if the subgroup K is not virtually abelian then for some $\varepsilon < 1/3$ there is an ε-embedding of $\mathbb{T}$ into K, relative to the metric ρ inherited from $\mathbb{U}_D$, such that for every vertex v of $\mathbb{T}$ at depth n, the image $f(v)$ has word length $\leq Cn$ in the generators of K, for some finite constant C. This will contradict the hypothesis that K is of polynomial growth, because Lemma 15.2.6 will imply that the number of distinct group elements $f(v)$ among vertices at depth $\leq n$ is at least 2^n.

The key to constructing the embedding is the fact that for any two elements $g, h \in \mathbb{U}_D$ within a sufficiently small distance $\varepsilon > 0$ of the group identity I (in the invariant metric ρ defined by equation (15.2.4)), the *commutator*

$$[g, h] := ghg^{-1}h^{-1} \tag{15.2.6}$$

is even closer to I. In particular, by the invariance of the metric ρ,

$$\begin{aligned}\rho(I, [g, h]) &= \rho(hg, gh)\\ &= \|hg - gh\|\\ &= \|(I-h)(I-g) - (I-g)(I-h)\|\\ &\leq 2\,\|I-h\|\;\|I-g\| = 2\rho(I, g)\rho(I, h).\end{aligned} \tag{15.2.7}$$

Obviously, this will only be useful for group elements g, h that are not too far from I. For this reason, we define, for any $\varepsilon > 0$,

$$K_\varepsilon := \langle K \cap \{g \in \mathbb{U}_D \; : \; \rho(I, g) < \varepsilon\}\rangle \tag{15.2.8}$$

to be the subgroup of K generated by all elements at distance $< \varepsilon$ from the identity.

Lemma 15.2.8 *The subgroup K_ε is a normal subgroup, has finite index in K, and has a finite set of generators all within distance ε of the group identity I.*

Proof. That K_ε has finite index in K follows from the compactness of the unitary group $\mathbb{U}_D$. The argument is as follows. By compactness, there exists a finite subset $\{x_i\}_{i \le m} \subset K$ such that every element of K is within distance ε of some x_i. Let $y \in K$ be any element of K; then for some element x_i we have $\rho(y^{-1}, x_i) < \varepsilon$, and so by invariance of the metric, $\rho(I, yx_i) < \varepsilon$. But this implies that $yx_i \in K_\varepsilon$, so y lies in the right coset $K_\varepsilon x_i^{-1}$. Thus, the (not necessarily distinct) cosets $K_\varepsilon x_i^{-1}$ cover K.

Next, we will show that K_ε is finitely generated. Let $\mathbb{A}$ be a finite, symmetric generating set for K. Since $\mathbb{U}_D$ is compact, there is a finite, $(\varepsilon/2)$-dense subset $\{y_i\}_{i \le m}$ of K (that is, for every $x \in K$ there is at least one y_i such that $\rho(x, y_i) < \varepsilon/2$). Define W_m to be the set of all elements of K of word-length $\le m$ relative to the generating set $\mathbb{A}$, and let $\mathbb{A}^*$ be the set of all $g \in K$ of the form

$$g = vwv^{-1} \quad \text{where} \;\; v \in W_m, \quad w \in W_m, \quad \text{and} \;\; \rho(I, w) < \varepsilon. \tag{15.2.9}$$

The set $\mathbb{A}^*$ is obviously finite, because $\mathbb{A}$ is finite, and it is symmetric, because $\mathbb{A}$ is symmetric and the metric ρ, being invariant, satisfies $\rho(I, w) = \rho(I, w^{-1})$. Furthermore, the invariance of ρ implies that for any $g \in K$ of the form (15.2.9),

$$\rho(I, g) = \rho(I, vwv^{-1}) = \rho(I, w) < \varepsilon;$$

thus, every element of $\mathbb{A}^*$ is within ρ-distance ε of I. □

Claim 15.2.9 *$\mathbb{A}^*$ is a generating set for K_ε.*

Proof of the Claim. Since $\mathbb{A}$ generates K, every element $x \in K_\varepsilon$ can be represented as a word $x = a_{i_1} a_{i_2} \cdots a_{i_n}$ in elements of $\mathbb{A}$. Now in every set of $m + 1$ elements of K there must be at least two distinct elements z_1, z_2 both within ρ-distance $\varepsilon/2$ of the same y_i, by the pigeonhole principle; since the metric ρ is invariant, any two such elements z_1, z_2 satisfy $\rho(I, z_1^{-1} z_2) < \varepsilon$. Therefore, the word $x = a_{i_1} a_{i_2} \cdots a_{i_n}$ can be broken into sub-words

$$x = \alpha w_1 w_2 \cdots w_l \beta$$

where $\alpha, \beta, w_i \in W_m$ and each w_i is within distance ε of I. But this shows that x can be represented as a finite word in the elements of $\mathbb{A}^*$:

$$x = (\alpha w_1 \alpha^{-1})(\alpha w_2 \alpha^{-1}) \cdots (\alpha w_l \beta).$$

$\square$

It remains to show that K_ε is a normal subgroup, i.e., that for every $x \in K$ the conjugate subgroup $xK_\varepsilon x^{-1}$ equals K_ε. Since $\mathbb{A}^*$ is a generating set for K_ε, it is enough to show that for every $g \in \mathbb{A}^*$ the conjugate xgx^{-1} is an element of K_ε, and for this it suffices to show that $\rho(xgx^{-1}, I) < \varepsilon$. By definition, every element $y \in \mathbb{A}^*$ is of the form $g = vwv^{-1}$, where $\rho(I, w) < \varepsilon$. Hence, by the invariance of the metric ρ,

$$\rho(xgx^{-1}, I) = \rho((xv)w(xv)^{-1}, I) = \rho(w, I) < \varepsilon.$$

Definition 15.2.10 For any group G, the *center* of G is the subgroup

$$Z(G) = \{h \in G \, : \, hg = gh \;\text{ for all } g \in G\}\,.$$

Any scalar multiple λI of the identity matrix I such that $|\lambda| = 1$ is in the center of $\mathbb{U}_D$, and therefore in the center of any subgroup of $\mathbb{U}_D$. The set $\{\lambda I\}_{|\lambda|=1}$ of all unitary matrices that are scalar multiples of I is obviously closed under multiplication, and therefore is a subgroup of $\mathbb{U}_D$.

Lemma 15.2.11 *If K is a finitely generated subgroup of $\mathbb{U}_D$ with at most polynomial growth, then for all sufficiently small $\varepsilon > 0$ either $K_\varepsilon \subset \{\lambda I\}_{|\lambda|=1}$ or the center $Z(K_\varepsilon)$ of K_ε contains an element not in $\{\lambda I\}_{|\lambda|=1}$.*

Proof. Suppose that $K_\varepsilon \not\subset \{\lambda I\}_{|\lambda|=1}$ but that $Z(K_\varepsilon) \subset \{\lambda I\}_{|\lambda|=1}$. By Lemma 15.2.8, there is a finite, symmetric generating set $\mathbb{A}^* = \{g_i\}_{i \le s}$ for K_ε consisting of elements g_i at ρ-distance $< \varepsilon$ from the identity I. Since $K_\varepsilon \not\subset \{\lambda I\}_{|\lambda|=1}$, at least one of the generators, which we may designate g_1, is not a scalar multiple of I. Moreover, for every $h \in K_\varepsilon \setminus \{\lambda I\}_{|\lambda|=1}$ there exists at least one generator g_i such that h does not commute with g_i, because otherwise h would be in the center of K_ε, which by hypothesis is contained in $\{\lambda I\}_{|\lambda|=1}$.

Define a sequence $(h_n)_{n \ge 0}$ in $K_\varepsilon \setminus Z(K_\varepsilon)$ inductively as follows. First, set $h_0 = g_1$. Then, assuming $h_0, h_1, h_2, \cdots, h_n$ that have been defined, let $g_{i_n} \in \mathbb{A}^*$ be a generator such that $[g_{i_n}, h_n] \neq I$; such a generator exists because by the induction hypothesis $h_n \notin Z(K_\varepsilon)$. Now define

$$h_{n+1} = [g_{i_n}, h_n] = g_{i_n} h_n g_{i_n}^{-1} h_n^{-1}.$$

By the product rule for determinants, $\det h_{n+1} = 1$ for every n; hence, if $h_{n+1} = \lambda I$ were a scalar multiple of the identity, then λ would be a Dth root of unity. By definition, $h_{n+1} \neq I$, so if $h_{n+1} = \lambda I$ then

$$\rho(h_{n+1}, I) = \min_{1 \le j < D} |e^{2\pi i j/D} - 1| > 0.$$

But by the commutator distance inequality (15.2.7) and the assumption that $\rho(I, g_i) < \varepsilon$ for every generator $g_i \in \mathbb{A}^*$,

$$0 < \rho(I, h_{n+1}) \leq 2\varepsilon\rho(I, h_n) \leq (2\varepsilon)^n \varepsilon. \tag{15.2.10}$$

Therefore, if $\varepsilon > 0$ is sufficiently small then no element of the sequence $(h_n)_{n\geq 0}$ is contained in $\{\lambda I\}_{|\lambda|=1}$, and hence no h_n is in the center of K_ε. If h_n has word length m in the generators g_i, then h_{n+1} has word length $\leq 2m+2$; consequently, the word length of h_n grows at most linearly with n.

Define a mapping f from the infinite binary tree to K_ε by setting $f(\emptyset) = h_0$ and

$$f(x_1x_2\cdots x_n) = h_1^{x_1}h_2^{x_2}\cdots h_n^{x_n} \quad \text{for any} \;\; x_1x_2\cdots x_n \in \{0, 1\}^n .$$

Since the word length of h_n grows at most linearly with n, the word length of $f(w)$ for any binary sequence of length n is at most Cn^2, for some $C < \infty$ independent of n. Inequality (15.2.10) and the invariance of the metric ρ imply that f is a (2ε)-tree embedding, because invariance implies that for any binary sequence $x_1x_2\cdots x_n$,

$$\rho(f(x_1x_2\cdots x_n), f(x_1x_2\cdots x_n1)) = \rho(I, h_{n+1}) \leq 2\varepsilon\rho(I, h_n).$$

Therefore, if $2\varepsilon < 1/3$ then we have a contradiction to the hypothesis that K (and hence also K_ε) has polynomial growth, because Lemma 15.2.6 implies that the number of distinct group elements in the set $\{f(v) : \text{depth}(v) \leq n\}$ is 2^n, but each such group element has word length $\leq Cn^2$. This proves that if the group K_ε is not contained in the set $\{\lambda I\}_{|\lambda|=1}$ then the center $Z(K_\varepsilon)$ is not contained in $\{\lambda I\}_{|\lambda|=1}$. □

Proof of Theorem 15.2.1. We will show that for all sufficiently small $\varepsilon > 0$ the group K_ε is virtually abelian. Since K_ε is a finite-index subgroup of K, by Lemma 15.2.8, this will imply that K is also virtually abelian.

The proof will proceed by induction on the dimension D. Since the unitary group $\mathbb{U}_1$ is abelian, any subgroup K of $\mathbb{U}_1$ is abelian, and so in the case $D = 1$ there is nothing to prove. Assume, then, that $D > 1$ and that the theorem is true for all dimensions $< D$.

By Lemma 15.2.11, either $K_\varepsilon \subset \{\lambda I\}_{|\lambda|=1}$, in which case K_ε is abelian, or the center of K_ε contains an element g that is not a scalar multiple of the identity matrix. By the Spectral Theorem, any such element g has a complete orthonormal basis of eigenvectors; at least two of the associated eigenvalues λ_i must be distinct, because otherwise g would be a scalar multiple of I. Let $\lambda_1, \lambda_2, \cdots, \lambda_D$ be an enumeration of the eigenvalues of g, and define orthogonal vector subspaces

$$V := \left\{ v \in \mathbb{C}^D \; : \; gv = \lambda_1 v \right\} \quad \text{and}$$

$$W := \left\{ w \in \mathbb{C}^D \; : \; \langle w, v \rangle = 0 \;\; \text{for all } v \in V \right\}.$$

Since the eigenspaces of distinct eigenvalues are orthogonal, by the Spectral Theorem, the subspace W contains all of the eigenspaces associated with eigenvalues $\neq \lambda_1$; hence, since g has at least one eigenvalue distinct from λ_1, each of the subspaces V, W has dimension ≥ 1. Moreover, since V, W are complementary, their dimensions sum to D. Thus, each of the subspaces V, W has dimension $< D$. These subspaces are both invariant by g (that is, $gV = V$ and $gW = W$); since g is in the center of K_ε, this has the following consequence. □

Claim 15.2.12 *The subspaces V and W are invariant by each $h \in K_\varepsilon$.*

Exercise 15.2.13 Prove this.

Consequently, for each $h \in K_\varepsilon$, the restrictions $h \restriction V$ and $h \restriction W$ are unitary linear transformations of V and W, respectively, relative to the inherited inner product (15.2.1). This implies that

$$K_\varepsilon^V = \{h \restriction V \,:\, h \in K_\varepsilon\} \quad \text{and}$$
$$K_\varepsilon^W = \{h \restriction W \,:\, h \in K_\varepsilon\}$$

are subgroups of unitary groups $\mathbb{U}_{\dim(V)}$ and $\mathbb{U}_{\dim(W)}$ of dimensions $< D$. The natural projections $\psi_V : K_\varepsilon \to K_\varepsilon^V$ and $\psi_W : K_\varepsilon \to K_\varepsilon^W$ are surjective — although not necessarily injective — homomorphisms, so the groups K_ε^V and K_ε^W are finitely generated groups of polynomial growth. Thus, by the induction hypothesis, K_ε^V is virtually abelian: in particular, it has a finite-index abelian subgroup H_ε^V. Since the homomorphism $\psi_V : K_\varepsilon \to K_\varepsilon^V$ is surjective, it follows that $H_\varepsilon := \psi_V^{-1}(H_\varepsilon^V)$ is a finite-index subgroup of K_ε, and hence finitely generated with at most polynomial growth. Consequently, the image $\psi_W(H_\varepsilon)$ in K_ε^W is a finitely generated subgroup of $\mathbb{U}_{\dim(W)}$ with at most polynomial growth, so by the induction hypothesis it has an abelian subgroup H_ε^W of finite index. The inverse image $G_\varepsilon := \psi_W^{-1}(H_\varepsilon^W)$ is a finite-index subgroup of H_ε, and hence also a finite-index subgroup of K_ε. Therefore, to complete the proof of Theorem 15.2.1 it will suffice to establish the following claim.

Claim 15.2.14 *The group G_ε is abelian.*

Proof. It suffices to prove that for any two elements $g_1, g_2 \in G_\varepsilon$ and any vector $z \in \mathbb{C}^D$,

$$g_1(g_2 z) = g_2(g_1 z). \tag{15.2.11}$$

Since the subspaces V, W are orthogonal complements, every vector $z \in \mathbb{C}^D$ can be decomposed as a sum $z = v + w$, where $v \in V$ and $w \in W$; consequently, it suffices to prove (15.2.11) for vectors $z = v \in V$ and $z = w \in W$. By definition of ψ_V, ψ_W,

$$g_1(g_2 v) = \psi_V(g_1)(\psi_V(g_2)v) \quad \text{for all } v \in V,$$

$$g_1(g_2 w) = \psi_W(g_1)(\psi_W(g_2)w) \quad \text{for all } w \in W.$$

By construction, the linear transformations $\psi_V(g_1)$ and $\psi_V(g_2)$ are elements of H_ε^V, and the linear transformations $\psi_W(g_1)$ and $\psi_W(g_2)$ are elements of H_ε^W. These groups are both abelian, so

$$\psi_V(g_1)(\psi_V(g_2)v) = \psi_V(g_2)(\psi_V(g_1)v) = g_2(g_1 v) \quad \text{for all } v \in V,$$
$$\psi_W(g_1)(\psi_W(g_2)w) = \psi_W(g_2)(\psi_W(g_1)w) = g_2(g_1 w) \quad \text{for all } w \in W.$$

Thus, (15.2.11) holds for all vectors $z = v \in V$ and $z = w \in W$. □

15.3 Milnor's Lemma

Proposition 15.1.6 and Corollary 15.2.2 imply that, regardless of whether the Kleiner image $\mathcal{L}^0(\Gamma)$ is finite or infinite, the group Γ has a finite-index subgroup Γ_0 which admits a surjective homomorphism φ onto some integer lattice $\mathbb{Z}^k$. If this homomorphism were an *isomorphism*, then Γ would be virtually abelian. In general, however, the homomorphism φ will have a nontrivial kernel. For such cases, the following result of Milnor [97] points the way to unravelling the structure of Γ.

Lemma 15.3.1 (Milnor) *Let Γ be a finitely generated group with polynomial growth of degree d. Suppose that there is a surjective homomorphism $\varphi : \Gamma \to \mathbb{Z}$ with kernel $ker(\varphi) = K$. Then K is finitely generated and has polynomial growth of degree at most $d - 1$.*

Exercise 15.3.2 Show that if $\varphi : \Gamma \to \mathbb{Z}$ is a surjective homomorphism with kernel K and if $g \in \Gamma$ is an element such that $\varphi(g) = 1$ then every $x \in \Gamma$ has a *unique* representation $x = g^n k$ where $k \in K$ and $n \in \mathbb{Z}$.

Proof of Lemma 15.3.1. Choose a generating set

$$\mathbb{A} = \left\{ g^{\pm 1}, e_1^{\pm 1}, e_2^{\pm 1}, \cdots, e_k^{\pm 1} \right\}$$

for Γ such that $\varphi(g) = 1$ and $\varphi(e_i) = 0$ for each $i \in [k]$. That such a generating set exists can be seen as follows. Start with any generating set $\mathbb{A}_0 = \left\{ f_i^{\pm 1} \right\}$, and then add the elements $g^{\pm 1}$, where g is any group element such that $\varphi(g) = 1$. Then replace each element f_i of $\mathbb{A}_0$ by $e_i := g^{-n} f_i$, where $n = n_i = \varphi(f_i)$. The resulting set $\mathbb{A} = \left\{ g^{\pm 1}, e_1^{\pm 1}, e_2^{\pm 1}, \cdots, e_k^{\pm 1} \right\}$ is a generating set with the desired properties.

We will build a finite generating set for K using the elements of $\mathbb{A}$ as building blocks. For each $i \in [k]$ and $m \in \mathbb{Z}$, define $z_{m,i} = g^m e_i g^{-m}$. □

Exercise 15.3.3 Show that the set $\left\{z_{m,i}^{\pm 1}\right\}_{i\in[k],m\in\mathbb{Z}}$ is a generating set for K, that is, every element $y \in K$ can be written as a finite product $y = x_1 x_2 \cdots x_n$ with each factor $x_j \in \left\{z_{m,i}^{\pm 1}\right\}_{i\in[k],m\in\mathbb{Z}}$.

The generating set $\left\{z_{m,i}^{\pm 1}\right\}_{i\in[k],m\in\mathbb{Z}}$ is, of course infinite; our goal is to build a *finite* generating set for K. To this end, for each index $i \in [k]$ and each $m \in \mathbb{N}$, let

$$G_{m,i}^{+} = \left\{z_{0,i}^{\varepsilon_0} z_{1,i}^{\varepsilon_1} \cdots z_{m,i}^{\varepsilon_m} \,:\, (\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_m) \in \{0,1\}^{m+1}\right\} \quad \text{and}$$

$$G_{m,i}^{-} = \left\{z_{0,i}^{\varepsilon_0} z_{-1,i}^{\varepsilon_1} \cdots z_{-m,i}^{\varepsilon_m} \,:\, (\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_m) \in \{0,1\}^{m+1}\right\}.$$

Clearly, the sets $G_{m,i}^{\pm}$ are nested, that is, $G_{m,i}^{\pm} \subset G_{m+1,i}^{\pm}$; moreover, each element of $G_{m,i}^{\pm}$ has word length at most $(m+1)(m+3)$ relative to the generating set $\mathbb{A}$. Since the cardinalities 2^{m+1} of the sets $\{0,1\}^{m+1}$ grow exponentially, the polynomial growth hypothesis implies that for all large m at least two words in each of the sets $G_{m,i}^{\pm}$ represent the same group element. Thus, there are 0–1 sequences ε, δ such that for some (maximal) $n \leq m$, $\varepsilon_n = 0$ and $\delta_n = 1$, and

$$z_{0,i}^{\varepsilon_0} z_{1,i}^{\varepsilon_1} \cdots z_{n,i}^{\varepsilon_n} = z_{0,i}^{\delta_0} z_{1,i}^{\delta_1} \cdots z_{n,i}^{\delta_n}$$

This implies that $z_{n,i}$ is an element of the group generated by the set $G_{n-1,i}^{+} \cup G_{n-1,i}^{-}$. The same argument shows that for some $n' \leq m$ the element $z_{-n',i}$ is in the group generated by $G_{n-1,i}^{+} \cup G_{n'-1,i}^{-}$. Hence, since the sets $G_{m,i}^{\pm}$ are nested, it follows that for some $n(i) \in \mathbb{N}$,

$$z_{m,i}^{\pm 1} \in \text{group generated by } G_{n(i),i}^{+} \cup G_{n(i),i}^{-} \quad \text{for all } \; |m| \geq n(i),$$

and so the finite set $\cup_{i\in[k]} G_{n(i),i}^{+} \cup G_{n(i),i}^{-}$ generates K.

Now let's show that K has polynomial growth of degree $d-1$. Recall (cf. Exercise 14.1.5) that the polynomial growth rate of a finitely generated group does not depend on the choice of generating set; hence, we can assume that word lengths in the group Γ are computed using the word length norm for the generating set $\mathbb{A} := \mathbb{A}_K \cup \left\{g, g^{-1}\right\}$, where $\mathbb{A}_K$ is a finite, symmetric generating set for K and $g \in \Gamma$ a group element such that $\varphi(g) = 1$. Let B_N^K be the ball of radius N in the group K, that is, the set of all group elements that can be represented as words in the alphabet $\mathbb{A}_K$ of length $\leq N$. For each fixed $N \in \mathbb{N}$, the left translates $g^j B_N^K$ are pairwise disjoint (because φ maps every element of $g^j B_N^K$ to j), and all have the same cardinality $|g^j B_N^K| = |B_N^K|$. Moreover, each element $g^j x \in g^j B_N^K$ has norm (relative to the generating set $\mathbb{A}$) $\leq |j| + N$. Hence, the set $\cup_{|j|\leq N} g^j B_N^K$ is contained in the ball $\mathbb{B}_{2N}$ in Γ. This implies that

$$N|B_N^K| \leq |\mathbb{B}_{2N}| \quad \text{for every } N \geq 1. \tag{15.3.1}$$

By the polynomial growth hypothesis, for any $\Delta > d$ there are infinitely many integers $N \geq 1$ for which the ball $\mathbb{B}_{2N}$ has cardinality $\leq C(2N)^{\Delta}$; therefore, K has polynomial growth of degree at most $d - 1$.

Proposition 15.1.6 and Corollary 15.2.2 imply that for an infinite, finitely generated group Γ with polynomial growth of degree d there exists a finite-index subgroup Γ_0 that admits a surjective homomorphism $\varphi_0 : \Gamma_0 \to \mathbb{Z}$; Milnor's Lemma implies that the kernel $K_0 := \ker(\varphi_0)$ is a finitely generated group with polynomial growth of degree at most $d - 1$. Thus, since the same logic applies to K_0, we have an inductive means of constructing finitely generated subgroups of decreasing degrees. The induction must terminate with a finite subgroup in no more than d steps, because there are no infinite groups with polynomial growth of degree < 1. Let's summarize this as follows.

Corollary 15.3.4 *For any infinite, finitely generated group Γ with polynomial growth of degree d there is, for some integer $0 \leq r \leq d$, a chain of finitely generated subgroups*

$$\Gamma \supset \Gamma_0 \supset K_0 \supset \Gamma_1 \supset K_1 \supset \cdots \Gamma_r \supset K_r \tag{15.3.2}$$

such that Γ_0 is a finite-index subgroup of Γ and for each $i \geq 0$,

(C1) Γ_{i+1} is a finite-index subgroup of K_i;
(C2) there is a surjective homomorphism $\varphi_i : \Gamma_i \to \mathbb{Z}$;
(C3) $K_i = ker(\varphi_i)$; and
(C4) K_r is finite.

15.4 Sub-cubic Polynomial Growth

We now have the tools necessary to tie up the loose end in the proof of Varopoulos' Recurrence Criterion (Section 7.3).

Theorem 15.4.1 *An infinite, finitely generated group either has a finite-index subgroup isomorphic to $\mathbb{Z}$ or $\mathbb{Z}^2$ or it has at least cubic growth, that is, there exists a constant $\kappa > 0$ such that for every $m \in \mathbb{N}$ the ball $\mathbb{B}_m$ of radius m in the Cayley graph has volume*

$$|\mathbb{B}_m| \geq \kappa m^3. \tag{15.4.1}$$

The proof will occupy the remainder of this section. We may assume that the group Γ has polynomial growth, because otherwise (15.4.1) must hold for all but finitely many m, by definition. Let's begin with two exercises in elementary group theory.

Exercise 15.4.2 If G is a group with a subgroup H of finite index $m = [G : H]$, then for $n = m!$ the subgroup $G^n := \langle g^n : g \in G \rangle$ generated by all nth powers of elements of G is a *characteristic* subgroup of G that is contained in H.

NOTE: A *characteristic subgroup* of a group G is a subgroup that is preserved by every automorphism of G. Every characteristic subgroup is normal, because for any group element $x \in G$ the mapping $g \mapsto xgx^{-1}$ is an automorphism of G.

HINT: Every element $g \in G$ induces a permutation π_g of the left cosets $\{x_i H\}_{i \in [m]}$ by left multiplication, that is, $\pi_g(x_i H) = (gx_i)H$.

Exercise 15.4.3 Let G be a group with a normal subgroup K. If K has finite cardinality $r = |K| < \infty$ then the subgroup $G^{r!}$ is contained in the centralizer of K, that is, for any $g \in G$ and $k \in K$,

$$g^{r!}k = kg^{r!}. \tag{15.4.2}$$

Proof of Theorem 15.4.1. Corollary 15.3.4 implies that there is a finite chain (15.3.2) of finitely generated subgroups such that the stipulations (C1)–(C4) all hold. Let $r \geq 0$ be the length of this chain. We will show that

(A) if $r = 0$ then Γ has a finite-index subgroup isomorphic to $\mathbb{Z}$;
(B) if $r = 1$ then Γ has a finite-index subgroup isomorphic to $\mathbb{Z}^2$; and
(C) if $r \geq 2$ then Γ has at least cubic growth.

□

The Case $r = 0$. In this case, Γ has a finite-index subgroup Γ_0 that has a homomorphic projection $\varphi_0 : \Gamma_0 \to \mathbb{Z}$ with finite kernel K_0. Let $g \in \Gamma_0$ be an element such that $\varphi_0(g) = 1$; then g has infinite order, because $\varphi(g^n) = n$ for every integer n, and so the cyclic subgroup $H := \{g^n\}_{n \in \mathbb{Z}}$ generated by g is isomorphic to $\mathbb{Z}$. This subgroup has finite index in Γ_0, because by Exercise 15.3.2,

$$\Gamma_0 = \bigcup_{k \in K_0} Hk. \tag{15.4.3}$$

Hence, since Γ_0 has finite index in Γ, the subgroup H has finite index in Γ.

The Case $r \geq 2$. Milnor's Lemma implies that for each pair of subgroups $\Gamma_i \supset K_i$ in the chain (15.3.2), the degree of polynomial growth of Γ_i exceeds that of K_i by at least one. Since $\Gamma_i \subset K_{i-1}$ for $i \geq 1$, the polynomial growth degree of K_{i-1} is at least that of Γ_i. Consequently, since Γ_2 has polynomial growth degree at least 1, the subgroup Γ_0 must have polynomial growth degree at least 3, and hence so must Γ.

The Case $r = 1$. In this case the subgroup chain (15.3.2) is

$$\Gamma \supset \Gamma_0 \supset K_0 \supset \Gamma_1 \supset K_1,$$

with the last subgroup K_1 finite. We will show that in this case there exist elements $h_i \in \Gamma_i$ for each $i = 0, 1$ such that $\varphi_i(h_i) \neq 0$ and such that $h_0 h_1 = h_1 h_0$. It will then follow that the subgroup $H = \langle h_0 . h_1 \rangle$ generated by h_0 and h_1 is isomorphic to $\mathbb{Z}^2$ and has finite index in Γ.

Set $r = [K_0 : \Gamma_1]!$, where $[K_0 : \Gamma_1]$ is the index of Γ_1 in K_0. By Exercise 15.4.2, the subgroup K_0^r is a characteristic subgroup of Γ_0, and $K_0^r \subset \Gamma_1$. Since the homomorphism $\varphi_1 : \Gamma_1 \to \mathbb{Z}$ is surjective, there is an element $g \in \Gamma_1$ such that $\varphi_1(g) = 1$. Thus, $\varphi_1(g^r) = r \neq 0$; since $g^r \in K_0^r$, it follows that K_0^r is not contained in $K_1 := \ker(\varphi_1)$. This implies that the restriction of φ_1 to K_0^r is a surjective homomorphism onto $d\mathbb{Z}$, for some $d \geq 1$ such that $d|r$. Hence, there exist elements $g_0 \in \Gamma_0$ and $g_1 \in K_0^r$ such that $\varphi_1(g_1) = d \neq 0$ and $\varphi_0(g_0) = 1$. Fix these elements and define

$$h_0 := g_0^2 \quad \text{and} \quad h_1 := g_1^{s!s} \quad \text{where} \quad s = |K_1|. \tag{15.4.4}$$

Let $H = \langle h_0, h_1 \rangle$ be the subgroup of Γ generated by h_0 and h_1.

Claim 15.4.4 *H is isomorphic to the additive group $\mathbb{Z}^2$.*

Claim 15.4.5 *H has finite index in Γ.*

Proof of Claim 15.4.4. By construction, the generators h_i satisfy $\varphi_i(h_i) \neq 0$, so both have infinite order in Γ. Therefore, to prove the Claim it suffices to show that

$$h_0 h_1 = h_1 h_0. \tag{15.4.5}$$

The subgroup $K_0 \subset \Gamma_0$ is normal, so it is invariant by conjugation with any element of Γ_0; in particular, the mappings $\alpha, \beta : K_0 \to K_0$ defined by

$$\begin{aligned} \alpha(x) &:= g_0 x g_0^{-1} \quad \text{and} \\ \beta(x) &:= g_1 x g_1^{-1} \end{aligned} \tag{15.4.6}$$

are group automorphisms. Since K_0^r is a characteristic subgroup of K_0, it follows that the restrictions of α, β to K_0^r are also automorphisms. For any $x \in K_0^r$,

$$\varphi_1(g_1 x g_1^{-1}) = \varphi_1(g_1) + \varphi_1(x) + \varphi_1(g_1^{-1}) = \varphi_1(x).$$

Therefore, the kernel $K_0^r \cap K_1$ of $\varphi_1 \restriction K_0^r$ is invariant by β. We will argue next that the same is true for *every* automorphism of K_0^r, and in particular α.

By Exercise 15.4.2, the group K_0^r is contained in Γ_1, and so the restriction of φ_1 to K_0^r is a homomorphism into $\mathbb{Z}$, with finite kernel $K_0^r \cap K_1$. By Exercise 15.3.2, each element $k \in K_0^r$ has a unique representation

$$k = g_1^n k_* \quad \text{where} \quad n \in \mathbb{Z} \quad \text{and} \quad k_* \in K_0^r \cap K_1. \tag{15.4.7}$$

By Lagrange's theorem for finite groups, every element of $K_0^r \cap K_1$ has finite order. On the other hand, each element $g_1^n k_*$, where $n \neq 0$ and $k_* \in K_0^r \cap K_1$, must be of *infinite order*, because the homomorphism φ_1 maps such elements to elements of infinite order in $\mathbb{Z}$; in particular,

$$\varphi_1(g_1^n k_*) = nd \neq 0.$$

Any automorphism of a group must send group elements to elements of the same order; therefore, the subgroup $K_0^r \cap K_1$, which consists of the finite-order elements of K_0^r, is invariant by any automorphism of K_0^r; in particular, the restrictions of α and β to $K_0^r \cap K_1$ are automorphisms.

Since α is an automorphism of K_0^r, the composition $\varphi_1 \circ \alpha : K_0^r \to \mathbb{Z}$ is a homomorphism with the same image $d\mathbb{Z}$ as φ_1. It follows that for some $k \in K_0^r \cap K_1$,

$$\alpha(g_1) = g_0 g_1 g_0^{-1} = g_1^{\pm 1} k. \tag{15.4.8}$$

To see this, observe that by (15.4.7) there exist $m \in \mathbb{Z}$ and $k' \in K_0^r \cap K_1$ such that $\alpha(g_1) = g_1^m k'$. Since $K_0^r \cap K_1$ is invariant by α, this implies that for any $n \in \mathbb{Z}$ and $k'' \in K_0^r \cap K_1$,

$$\varphi_1 \circ \alpha(g_1^n k'') = \varphi_1(g_1^{mn} k'') = \varphi(g_1^{mn}) = mnd;$$

thus, since every element of K_0^r is of the form $g_1^{nk''}$, the homomorphism $\varphi_1 \circ \alpha$ can have full image $d\mathbb{Z}$ only if $m = \pm 1$. Although this leaves ambiguity of the sign, it nevertheless follows by the invariance of $K_0^r \cap K_1$ under α that for some $k' \in K_0^r \cap K_1$.

$$\alpha^2(g_1) = g_0^2 g_1 g_0^{-2} = g_1 k'. \tag{15.4.9}$$

Any automorphism of a finite group is a *permutation* of its elements. The set of permutations of a finite set with m elements is itself a finite group of order $m!$; consequently, by Lagrange's theorem,

$$\beta^{s!}(k) = g_1^{s!} k g_1^{-s!} = k \quad \text{for every} \quad k \in K_0^r \cap K_1. \tag{15.4.10}$$

Equation (15.4.9) and the fact that $K_0^r \cap K_1$ is invariant by conjugation with g_1 imply that $\alpha^2(g_1^{s!}) = g_1^{s!} k_*$ for some $k_* \in K_0^r \cap K_1$, so by equation (15.4.10),

$$g_0^2 g_1^{s!s} g_0^{-2} = \alpha^2(g_1^{s!s}) = \alpha^2(g_1^{s!})^s = g_1^{s!s} k_*^s = g_1^{s!s}.$$

This proves (15.4.5), and hence Claim 15.4.4. $\square$

Proof of Claim 15.4.5. First, let's show that the cyclic subgroup $\langle h_1 \rangle$ generated by h_1 has finite index in K_0. Since the homomorphism $\varphi_1 : \Gamma_1 \to \mathbb{Z}$ is surjective, there exists $x_1 \in \Gamma_1$ such that $\varphi_1(x_1) = 1$, so by Exercise 15.3.2 every element $g \in \Gamma_1$ has a unique representation $g = x_1^n k$ for some $n \in \mathbb{Z}$ and $k \in K_1$. By construction, $\varphi_1(h_1) = s!sd \geq 1$, so $h_1^{-1} x_1^{s!s} \in K_1$. Consequently, every element $g \in \Gamma_1$ has a unique representation

$$g = h_1^n x_1^\ell k \quad \text{where } \; n \in \mathbb{Z}, k \in K_1, \quad \text{and } \; 0 \leq \ell < s!s.$$

This implies that the subgroup $\langle h_1 \rangle$ has finite index in Γ_1. Since Γ_1 has finite index in K_0, it follows that $\langle h_1 \rangle$ has finite index in K_0; in particular, there is a finite subset $\{y_i\}_{i \leq K} \subset K_0$ such that every element $x \in K_0$ has a unique representation

$$x = h_1^n y_i \quad \text{for some } \; n \in \mathbb{Z} \quad \text{and } \; i \leq K.$$

Next, we will use a similar argument to show that the subgroup $\langle h_0, h_1 \rangle$ has finite index in Γ_0. By Exercise 15.3.2, every element $x \in \Gamma_0$ has a unique representation $x = g_0^n k$, where $k \in K_0$, and hence, since K_0 is invariant by conjugation with g_0, a unique representation

$$x = h_0^m k' g_0^\ell \quad \text{where } \; m \in \mathbb{Z}, k' \in K_0, \quad \text{and } \; \ell \in \{0, 1\}.$$

But this implies that every element $x \in \Gamma_0$ has a unique representation

$$x = h_0^m h_1^n y_i k' g_0^\ell,$$

which implies that the subgroup $\langle h_0, h_1 \rangle$ has finite index in Γ_0. Since Γ_0 has finite index in Γ, the Claim follows.

□

15.5 Gromov's Theorem

Gromov's classification of finitely generated groups of polynomial growth is one of the signal achievements of the past 50 years in geometric group theory. In this section we will present Kleiner's proof of Gromov's main result, as simplified by Shalom & Tao.

Theorem 15.5.1 *(Gromov) If Γ is a finitely generated group of polynomial growth then Γ is* virtually nilpotent*, that is, Γ has a nilpotent subgroup of finite index.*

To unravel the meaning of this statement, let's review some standard terminology.

Definition 15.5.2 The *commutator* of two elements x, y of a group G is the element $[x, y] := xyx^{-1}y^{-1}$. The *commutator subgroup* $[H, K]$ generated by two

subgroups H, K of G is the smallest subgroup containing all commutators $[h, k]$, where $h \in H$ and $k \in K$. The *lower central series* of G is the sequence of subgroups $G = G^{(0)} \rhd G^{(1)} \rhd G^{(2)} \rhd \cdots$ defined inductively by $G^{(i+1)} = [G, G^{(i)}]$. The group G is *nilpotent* if for some $\nu \in \mathbb{N}$ the lower central series terminates in the trivial subgroup $G^{(\nu)} = \{1\}$; the smallest integer ν for which this occurs is the *nilpotence degree*.

Elementary Facts:

(N1) The subgroups $G^{(i)}$ are characteristic, and hence normal.
(N2) Each quotient group $G^{(i)}/G^{(i+1)}$ is abelian.
(N3) Subgroups of nilpotent groups are nilpotent.
(N4) Homomorphic images of nilpotent groups are nilpotent.

Exercise 15.5.3 Show that if G is nilpotent of degree ν then the quotient group $G/G^{(\nu-1)}$ is nilpotent of degree $\leq \nu - 1$.

Exercise 15.5.4 Let G be a group with normal subgroup N. Show that if the quotient group G/N is nilpotent of degree d then $G^{(d)} \subset N$.

Exercise 15.5.5 Show that if G is a finitely generated group of polynomial growth then so are its commutator subgroups $G^{(i)}$.

HINT: Let $\mathbb{A}$ be a finite generating set for G. Define finite sets $\mathbb{A}_i$ inductively as follows: let $\mathbb{A}_0 = \mathbb{A}$, and let $\mathbb{A}_{i+1}$ be the set of all commutators $[a, b]$ and $[b, a]$ where $a \in \mathbb{A}$ and $b \in \mathbb{A}_i$. Prove that $G^{(i)}$ is generated by $\mathbb{A}_i$. For this, the identity

$$ab = ba[a^{-1}, b^{-1}]$$

might be useful.

Example 15.5.6 † The Heisenberg group $G = \mathbb{H}$ is nilpotent of degree 2: the commutator subgroup $G^{(1)} = [G, G]$ is the subgroup of all matrices of the form

$$\begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{where} \quad c \in \mathbb{Z}.$$

More generally, the group $\mathbb{H}_m$ comprised of all upper triangular $m \times m$ matrices with integer entries and 1s on the main diagonal is nilpotent of degree $m - 1$.

For any finitely generated group G, the subgroups $G^{(i)}$ in the lower central series are characteristic, so they are fixed by conjugation with any element of G. More generally, let $g \in G$ and suppose that N is a subgroup of G fixed by the conjugation mapping

$$\alpha(x) = \alpha_g(x) := gxg^{-1} \quad \text{for} \quad x \in G. \tag{15.5.1}$$

If K is a normal subgroup of N that is also fixed by α then α induces an automorphism of the quotient group N/K. If in addition N/K is finitely generated and abelian (as are the quotients $G^{(i)}/G^{(i+1)}$ in the lower central series), then by the Structure Theorem for finitely generated abelian groups (cf. Herstein [65], Theorem 4.5.1), it is isomorphic to $\mathbb{Z}^r \oplus H$ for some $r \geq 1$, called the *rank* of N/K, where H is the subgroup of N/K consisting of all elements of finite order. This subgroup H, called the *torsion subgroup*, is finite. Any automorphism of N/K must preserve this splitting, and therefore restricts to an automorphism of the integer lattice $\mathbb{Z}^r$. A particular class of automorphisms, the *unipotent* automorphisms, will figure prominently in the proof of Gromov's Theorem.

Definition 15.5.7 An automorphism $\alpha : \mathbb{Z}^r \to \mathbb{Z}^r$ is *unipotent* if

$$(I - \alpha)^r = 0. \tag{15.5.2}$$

An automorphism α of a finitely generated abelian group is unipotent if its restriction to its torsion-free subgroup $\mathbb{Z}^r$ is unipotent.

Exercise 15.5.8 Show that if $\alpha : G \to G$ is a unipotent automorphism of a finitely generated abelian group then so is any iterate

$$\alpha^n = \alpha \circ \alpha \circ \cdots \circ \alpha \quad (n \text{ times}).$$

Exercise 15.5.9 Suppose that $\alpha : \mathbb{Z}^r \to \mathbb{Z}^r$ is a unipotent automorphism, and let $V = \{x \in \mathbb{Z}^r : \alpha(x) = x\}$ be the subgroup consisting of all vectors fixed by α.

(A) Show that there exists a nonzero vector $x \in V$.
(B) Show that the quotient group $\mathbb{Z}^r/V$ has trivial torsion subgroup, that is, $\mathbb{Z}^r/V$ has no nonzero elements of finite order.
(C) Show that α induces a *unipotent* automorphism of $\mathbb{Z}^r/V$.

An automorphism $\alpha : \mathbb{Z}^r \to \mathbb{Z}^r$ uniquely determines — and is uniquely determined by — an $r \times r$ matrix with integer entries. Because the mapping α is bijective, the corresponding matrix must have determinant ± 1. Thus, α must either have an eigenvalue of absolute value > 1, or all of its eigenvalues must have absolute value 1.

Proposition 15.5.10 (Kronecker) *If an automorphism $\alpha : \mathbb{Z}^r \to \mathbb{Z}^r$ has no eigenvalue of absolute value greater than* 1 *then for some $m \in \mathbb{N}$ the mth iterate α^m of α is unipotent.*

Proof. The eigenvalues of α are the roots of its characteristic polynomial $f(z) := \det(zI - \alpha)$. Since the matrix entries of α are integers, so are the coefficients of $f(z)$, and the coefficient of the leading term z^r is 1. Denote by $\zeta_1, \zeta_2, \cdots, \zeta_r$ the roots of $f(z)$, listed according to multiplicity; then

$$f(z) = \prod_{i=1}^{r}(z - \zeta_i) = z^r + \sum_{k=1}^{r} z^{r-k} s_k(\zeta_1, \zeta_2, \cdots, \zeta_r), \tag{15.5.3}$$

where $s_1, s_2, \cdots$ are the elementary symmetric polynomials

$$s_1 = \sum_i \zeta_i, \quad s_2 = \sum_{i<j} \zeta_i \zeta_j, \quad \cdots, \quad s_r = \zeta_1 \zeta_2 \cdots \zeta_r \tag{15.5.4}$$

(cf. Herstein [65], Section 5.6). Since $f(z)$ is a monic polynomial with integer coefficients, each $s_i \in \mathbb{Z}$.

The representation (15.5.3) holds for any monic polynomial with integer coefficients. If the roots of such a polynomial all have absolute value 1, then each symmetric polynomial $s_i = s_i(\zeta_1, \zeta_2, \cdots, \zeta_r)$ is bounded in absolute value, in particular, $|s_j| \leq \binom{r}{j}$. Therefore, there are only finitely many monic polynomials with integer coefficients whose roots all lie on the unit circle. Denote the set of all such polynomials by $\mathcal{S}_r$.

For any $n \in \mathbb{N}$ the nth iterate α^n of α is an automorphism of $\mathbb{Z}^r$ whose eigenvalues are $\zeta_1^n, \zeta_2^n, \cdots, \zeta_r^n$; these have the same multiplicities as for the automorphism α. (See Exercise 15.5.11 below.) Consequently, $\det(z - \alpha^n) = \prod_{i=1}^{r}(z - \zeta_i^n) := f_n(z)$, and so the polynomial $f_n(z)$ is monic and has integer coefficients. The roots ζ_i^n all have absolute value 1, so they must all belong to the finite set consisting of all roots of polynomials in $\mathcal{S}_r$. This implies that for each eigenvalue ζ_i, the cyclic group $\{\zeta_i^n\}_{n\in\mathbb{N}}$ generated by ζ_i is finite; thus, for each i there exists $m_i \in \mathbb{N}$ such that $\zeta_i^{m_i} = 1$.

Set $m = m_1 m_2 \cdots m_r$; then the eigenvalues ζ_i are all mth roots of unity. But $\zeta_1^m, \zeta_2^m, \cdots, \zeta_r^m$ are the eigenvalues of the automorphism α^m; therefore, the sole eigenvalue of α^m is 1, and so its characteristic polynomial must be a divisor of $\det(z - \alpha^m) = (z - 1)^r$. Since every linear transformation satisfies its own characteristic polynomial, this implies that α is unipotent.

□

Exercise 15.5.11 Use the *Jordan canonical form* of the matrix α (cf. Herstein [65], Chapter 6, Theorem 6.6.2) to show that for any linear transformation T of a finite-dimensional complex vector space and any $n \in \mathbb{N}$, the eigenvalues of T^n are the nth powers of the eigenvalues of T, with the same multiplicities.

The following proposition shows that if α is an automorphism of $\mathbb{Z}^r$ with an eigenvalue of absolute value greater than 1 then the matrix norms $\|\alpha^n\|$ of its powers grow exponentially with n.

Proposition 15.5.12 (Gelfand) *If T is a complex, $r \times r$ matrix with eigenvalues λ_i then*

$$\lim_{n\to\infty} \|T^n\|^{1/n} = \max_i |\lambda_i|. \tag{15.5.5}$$

This is a special case of a fundamental result in the theory of Banach algebras — see, for instance, Rudin [112], Theorem 18.9. The usual proof relies on properties of matrix-valued holomorphic functions. The following proof avoids the use of holomorphic function theory, using instead the existence of the Jordan canonical form.

Proof. The limit in (15.5.5) exists by submultiplicativity of the matrix norm. We may assume, without loss of generality, that the nonzero eigenvalues λ_i are all of absolute value greater than 1, because replacing T by a scalar multiple βT does not affect the validity of (15.5.5).

Any square matrix T with eigenvalues λ_i can be written in Jordan canonical form as $T = SBS^{-1}$, where B is a block diagonal matrix of the form

$$B = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{pmatrix} \quad \text{where} \quad J_i = \begin{pmatrix} \lambda_i & 1 & 0 & 0 \cdots & 0 \\ 0 & \lambda_i & 1 & 0 \cdots & 0 \\ & & \ddots & & \\ 0 & 0 & 0 & 0 \cdots & \lambda_i \end{pmatrix}.$$

Submultiplicativity of the matrix norm implies that for any $n \in \mathbb{N}$,

$$\left\| T^n \right\| \leq \|S\| \left\| S^{-1} \right\| \left\| B^n \right\| \quad \text{and}$$
$$\left\| B^n \right\| \leq \|S\| \left\| S^{-1} \right\| \left\| T^n \right\|,$$

so it suffices to prove (15.5.5) for matrices of the form $T = B$. But $\|B\| = \max_i \|J_i\|$, so it is enough to prove (15.5.5) for Jordan blocks J_i where either $\lambda_i = 0$ or $|\lambda_i| > 1$.

In the first case, where J_i has the sole eigenvalue $\lambda_i = 0$, the matrix J_i satisfies $J_i^{r_i} = 0$, where r_i is the dimension of J_i. This implies that $\left\| J_i^n \right\| = 0$ for every $n \geq r_i$, and so the equality (15.5.5) holds in this case.

In the second case, where $|\lambda_i| > 1$, a routine calculation shows that for any $n \geq 1$ the matrix J_i^n is upper triangular, with diagonal entries $(J_i^n)_{k,k} = \lambda_i^n$ and off-diagonal entries

$$(J_i^n)_{k,k+m} = \binom{n}{m} \lambda_i^{n-m}.$$

Since $|\lambda_i| > 1$, it follows that

$$\lim_{n \to \infty} \frac{J_i^n}{\lambda_i^n} = I,$$

and so the equality (15.5.5) holds for $T = J_i$. $\square$

We will now prove Gromov's Theorem 15.5.1 by induction on the degree of polynomial growth. Milnor's Lemma provides the catalyst for the induction: it implies that for any finitely generated group Γ with polynomial growth of degree d there is a finite-index subgroup Γ_0 that has a normal subgroup K_0 such that (i) the quotient group Γ_0/K_0 is isomorphic to $\mathbb{Z}$, and (ii) K_0 is itself finitely generated with polynomial growth of degree $\leq d-1$. The induction hypothesis will imply that K_0 has a finite-index *nilpotent* subgroup N. The nilpotence of N will allow us to build a finite-index nilpotent subgroup of Γ.

Recall that for any group G, the subgroup G^n is the smallest subgroup of G containing all nth powers g^n, where $g \in G$.

Lemma 15.5.13 *If N is a nilpotent subgroup of a finitely generated group G of finite index $m = [G : N]$ then for $n = m!$ the subgroup G^n is characteristic in G and has finite index.*

Proof. By Exercise 15.4.2, the subgroup G^n is characteristic and is contained in N, so G^n is a normal subgroup of N. Obviously, the order of any element of the quotient group N/G^n must divide n. Therefore, since the quotient N/G^n is nilpotent (cf. Fact (N4) above), the assertion is a consequence of the following fact about nilpotent groups. □

Lemma 15.5.14 *Any finitely generated, nilpotent group in which every element has finite order is finite.*

Proof. By induction on the degree of nilpotence. If the nilpotence degree is $\nu = 1$, then H is an abelian group; since every element has finite order, the Structure Theorem for finitely generated abelian groups implies that H is finite.

Suppose, then, that the assertion is true for all finitely generated, nilpotent groups of nilpotence degree $< \nu$, where $\nu \geq 2$. Let H be a finitely generated, nilpotent group of nilpotence degree ν, and let $H^{(\nu-1)}$ be the penultimate term in the lower central series. Then the quotient group $H/H^{(\nu-1)}$ is finitely generated and nilpotent, as it is a homomorphic image of H; it has nilpotence degree $\nu-1$; and every element has finite order. Consequently, by the induction hypothesis, $H/H^{(\nu-1)}$ is finite, and so $H^{(\nu-1)}$ has finite index in H. It follows that $H^{(\nu-1)}$ is finitely generated (cf. Exercise 15.5.5, or alternatively Exercise 1.2.11), and it is abelian, as $[H, H^{(\nu-1)}] = H^{(\nu)} = \{1\}$. Since every element of $H^{(\nu-1)}$ has finite order, the induction hypothesis implies that $H^{(\nu-1)}$ is finite. Therefore, H is finite. □

Let N be a finitely generated, normal subgroup of a group G and for any $i \geq 0$ denote by $N^{(i)}$ the ith term in the lower central series of N. Because each subgroup $N^{(i)}$ is characteristic in N, it is preserved by conjugation with any element $g \in G$; thus, if $\alpha = \alpha_g : G \to G$ is the conjugation mapping (15.5.1) then the restriction of α to any $N^{(i)}$ is an automorphism of $N^{(i)}$. Consequently, for each i the mapping α induces an automorphism, also denoted by α, of the quotient group $N^{(i)}/N^{(i+1)}$, by

$$xN^{(i+1)} \mapsto \alpha(x)N^{(i+1)}.$$

Each quotient group $N^{(i)}/N^{(i+1)}$ is a finitely generated abelian group (see Exercises 15.5.5 and 14.1.5).

Lemma 15.5.15 *Let N be a finitely generated, nilpotent, normal subgroup of a group G, and let $g \in G$ be an element for which the conjugation automorphism (15.5.1) induces unipotent automorphisms of each quotient group $N^{(i)}/N^{(i+1)}$. Then for some integer $m \geq 1$ the subgroup $\langle g^m, N\rangle$ generated by g^m and N is nilpotent.*

Proof. By induction on the nilpotence degree ν of N.

The Case $\nu = 1$. In this case the group N is abelian, and so $N \cong \mathbb{Z}^r \oplus H$ where H is finite. We proceed by induction on the rank r. Suppose that $r = 1$, and let $\alpha = \alpha_g$ be the automorphism (15.5.1). Since the restriction of α to H is an automorphism, its $|H|!$ iterate is the identity; thus, we can assume without loss of generality (replace g by $g^{|H|!}$; cf. Exercise 15.5.8) that α is the identity on H. Since the restriction of α to $\mathbb{Z}$ is unipotent, α is the identity map on $\mathbb{Z}$, and consequently also on N. This means that g commutes with every element of N, so the group $\langle g, N\rangle$ is abelian and hence nilpotent of degree 1.

Next, suppose that $N \cong \mathbb{Z}^r \oplus H$ is abelian of rank $r \geq 2$, with torsion subgroup H, and that the automorphism $\alpha = \alpha_g$ of N is unipotent. Once again we may assume that the restriction of α to the torsion subgroup H is the identity map. By Exercise 15.5.9, there exists a nonzero vector $e_r \in \mathbb{Z}^r$ that is fixed by α, and so α induces an automorphism of the quotient group $N/\langle e_r\rangle$. This quotient group is abelian of rank $r-1$; hence, by the induction hypothesis on r, there exists an integer $s \geq 1$ such that the group $\langle g^s, N\rangle / \langle e_r\rangle$ is nilpotent of degree $d \leq r-1$. This implies (cf. Exercise 15.5.4) that

$$\langle g, N\rangle^{(d)} \subset \langle e_r\rangle \quad \Longrightarrow \quad \langle g, N\rangle^{(d+1)} = \{1\}.$$

Therefore, the group $\langle g, N\rangle$ is nilpotent of degree $\leq r$. This proves the lemma in the case where N has nilpotence degree $\nu = 1$.

The Induction Step. Assume that the result holds for all finitely generated nilpotent groups of degree $< \nu$, for some $\nu \geq 2$, and let N be a finitely generated, nilpotent, normal subgroup of G of nilpotence degree ν. In this case $[N, N^{(\nu-1)}] = \{1\}$, so every element of N commutes with every element of $N^{(\nu-1)}$, and in particular $N^{(\nu-1)}$ is abelian. Both $N/N^{(\nu-1)}$ and $N^{(\nu-1)}$ are nilpotent of degree $< \nu$, so by the induction hypothesis there is an integer $m \geq 1$ such that the groups $\langle g^m, N\rangle / N^{(\nu-1)}$ and $\left\langle g^m, N^{(\nu-1)}\right\rangle$ are both nilpotent. Thus, by Exercise 15.5.4, for suitable integers $m, r, s \geq 1$,

$$\left\langle g^m, N\right\rangle^{(s)} \subset N^{(\nu-1)} \quad \text{and} \quad \left\langle g^m, N^{(\nu-1)}\right\rangle^{(r)} = \{1\}.$$

The conjugation operator $\alpha = \alpha_g$ fixes N, since N is a normal subgroup of G, so every element of the group $\langle g, N\rangle$ has the form $g^n x$ for some $x \in N$. Since elements

of N commute with elements of $N^{(\nu-1)}$, it follows that

$$[g^n x, z] = [g^n, z] \quad \text{for all} \;\; n \in \mathbb{Z}, x \in N, \quad \text{and} \;\; z \in N^{(\nu-1)}.$$

Therefore,

$$\begin{aligned}[\langle g^m, N\rangle, \langle g^m, N\rangle^{(s)}] &\subset [\langle g^m, N\rangle, N^{(\nu-1)}] \\ &= [\langle g^m\rangle, N^{(\nu-1)}] \\ &= \left\langle g^m, N^{(\nu-1)}\right\rangle^{(1)} \subset N^{(\nu-1)},\end{aligned}$$

the last inclusion by virtue of the fact that the subgroup $N^{(\nu-1)}$ is normal. Hence, by iteration,

$$\langle g^m, N\rangle^{(r+s)} \subset \left\langle g^m, N^{(\nu-1)}\right\rangle^{(r)} = \{1\}.$$

This proves that the group $\langle g^m, N\rangle$ is nilpotent of degree $r+s$. □

Lemma 15.5.16 *Let Γ be a finitely generated group of polynomial growth, and let $H \subset \Gamma$ be a finitely generated abelian subgroup. If $g \in \Gamma$ is an element such that the conjugation map $\alpha_g(x) = gxg^{-1}$ fixes H then there exists an integer $m \geq 1$ such that the restriction to H of the conjugation mapping $\alpha_{g^m} = \alpha_g^m$ is unipotent.*

Proof. By induction on the rank r of H. By hypothesis, H is a finitely generated abelian group, so $H \cong \mathbb{Z}^r \oplus T$ where r is the rank and T the torsion subgroup. The automorphism $\alpha = \alpha_g$ restricts to an automorphism of the torsion-free subgroup $\mathbb{Z}^r$, which we also denote by α. If $r = 1$ then α^2 must be the identity, because any automorphism of $\mathbb{Z}$ must be either the identity or the map $x \mapsto -x$.

For the induction step, it suffices, by Proposition 15.5.10, to show that the restriction $\alpha \restriction \mathbb{Z}^r$ has no eigenvalue of absolute value > 1. Suppose to the contrary that it does; we will show that this contradicts the hypothesis that Γ has polynomial growth.

Let $\rho = \max_i |\lambda_i| > 1$ be the modulus of the lead eigenvalue of the automorphism $\alpha \restriction \mathbb{Z}^r$. By Proposition 15.5.12,

$$\lim_{n\to\infty} \|\alpha^n\|^{1/n} = \rho;$$

consequently, for any $\varepsilon > 0$ there exist unit vectors $v_n \in \{e_i\}_{i\in[r]}$ such that for all sufficiently large $n \in \mathbb{N}$,

$$(1-\varepsilon)\rho^{n^2} \leq \left\|\alpha^{n^2} v_n\right\| \leq (1+\varepsilon)\rho^{n^2},$$

and so there exists $n_* \in \mathbb{N}$ such that for all $n \geq n_*$,

$$\left\| \alpha^{(n+1)^2} v_{n+1} \right\| \geq 4 \left\| \alpha^{n^2} v_n \right\| .$$

This implies (exercise!) that for any two distinct binary sequences $(b_k)_{n_* \leq k \leq n}$ and $(c_k)_{n_* \leq k \leq n}$ of length n,

$$\sum_{k=n_*}^{n} b_k \alpha^{k^2}(v_k) \neq \sum_{k=n_*}^{n} c_k \alpha^{k^2}(v_k).$$

But each such sum represents an element of Γ of word length $\leq \kappa n^3$, for some constant $\kappa < \infty$ depending on the generating set for the group Γ and the element g, because for each of the standard unit vectors e_i, the word length of $\alpha_g^m e_i$ is

$$|\alpha_g^m(e_i)| = |g^m e_i g^{-m}| \leq 2m|g| + |e_i|.$$

Thus, the ball of radius κn^2 in Γ contains at least 2^n distinct group elements, a contradiction to the polynomial growth hypothesis. □

Proof of Theorem 15.5.1. We will induct on the degree of polynomial growth. By Theorem 15.4.1, if a finitely generated group Γ has polynomial-growth degree less than 3 then it has a finite-index subgroup isomorphic to either $\mathbb{Z}$ or $\mathbb{Z}^2$, and therefore is virtually nilpotent. Assume, then, that the theorem is true for all finitely generated groups of polynomial-growth degree less than d, for some integer $d \geq 3$; we will show that it is also true for groups of polynomial-growth degree less than $d + 1$.

Let Γ be such a group. Proposition 15.1.6 and Corollary 15.2.2 imply that Γ has a finite-index subgroup Γ_0 that admits a surjective homomorphism $\varphi : \Gamma_0 \to \mathbb{Z}$, and Milnor's Lemma implies that the kernel K of this homomorphism is finitely generated and has polynomial-growth degree less than d. Thus, by the induction hypothesis, K has a finite-index, nilpotent subgroup N'. The subgroup N' is finitely generated, because it has finite index in K (Exercise 1.2.11).

Denote by $N = K^n$ the characteristic subgroup of K generated by all nth powers of elements in K, where $n = [K : N']!$. By Exercise 15.4.2 and Lemma 15.5.13, N is a finite-index subgroup of N', and hence also of K, and consequently is finitely generated. Moreover, since N' is nilpotent, so is N.

Let g be an element of Γ_0 such that $\varphi(g) = 1$, and let $\alpha = \alpha_g$ be the conjugation automorphism $\alpha(x) = gxg^{-1}$ associated with g. Clearly, α fixes the subgroup K; consequently, it also fixes N and each of its commutator subgroups N^i, as these are characteristic in K. Therefore, α induces automorphisms of the quotient groups $N/N^{(i+1)}$. These quotient groups are finitely generated, nilpotent groups of polynomial growth (cf. Exercises 14.1.5 and 15.5.5), and for each i the subgroup $N^{(i)}/N^{(i+1)}$ is finitely generated and abelian. Hence, by Lemma 15.5.16, for each i there is an integer $m_i \geq 1$ such that the restriction of α^{m_i} to $N^{(i)}/N^{(i+1)}$ is unipotent. Now set $m = \prod_i m_i$; since N is nilpotent, the product is finite, and

the automorphism α^m is unipotent on each quotient $N^{(i)}/N^{(i+1)}$. It now follows by Lemma 15.5.15 that the subgroup $\langle g^m, N\rangle$ of Γ generated by g^m and N is nilpotent.

It remains to check that the subgroup $\langle g^m, N\rangle$ has finite index in Γ. First, since $gN = Ng$, every element of $\langle g, N\rangle$ has a unique representation $g^n x$ for some $n \in \mathbb{Z}$ and $x \in N$; consequently, the subgroup $\langle g^m, N\rangle$ has finite index in $\langle g, N\rangle$. Second, since N is a characteristic subgroup of finite index in K, the automorphism $\alpha(x) = gxg^{-1}$ permutes cosets of N in K; consequently, $\langle g, N\rangle$ has finite index in $\langle g, K\rangle$. Finally, by Exercise 15.3.2,

$$\langle g, K\rangle = \Gamma_0,$$

which is a subgroup of finite index in Γ. Therefore, $\langle g^m, N\rangle$ has finite index in Γ. □

Additional Notes. The central result of this Chapter, Theorem 15.5.1, was proved by Gromov in his celebrated article [60]. The proof given in this Chapter is loosely modeled on that in Kleiner's article [80], following the modifications outlined in Tao's blog article [122]. An earlier result of Tits [125], the so-called *Tits Alternative*, implied the weaker result that every finitely generated subgroup of a connected Lie group either is virtually nilpotent or has exponential growth. Theorem 15.2.1 can be deduced from Tits' Theorem; the much more elementary proof given in Section 15.2 is adapted from Shalom & Tao [116, 122]. Theorem 15.4.1 is, of course, a special case of Gromov's Theorem; however, because Varopoulos' Theorem 7.3.1 uses only this special case, and because the proof is considerably simpler in this case, I have included it here.

Appendix A
A 57-Minute Course in Measure–Theoretic Probability

A.1 Definitions, Terminology, and Notation

A *probability space* is a measure space $(\Omega, \mathcal{F}, P)$ such that $P(\Omega) = 1$. Elements of the σ-algebra $\mathcal{F}$ are called *events*. A *random variable* is a measurable transformation $Y : \Omega \to \Upsilon$ with domain Ω; the range $(\Upsilon, \mathcal{G})$ can be any measurable space. Thus, if V is a countable set with σ-algebra 2^V then a V-valued random variable is a function $Y : \Omega \to V$ such that for every $v \in V$ the inverse image $Y^{-1}\{v\}$ is an element of $\mathcal{F}$. Random variables are customarily (but not always) denoted by upper case letters U, V, W, X, Y, Z near the end of the alphabet, or the lower case Greek letters ξ, ζ, τ. For any event F, the *indicator* of F is the random variable $\mathbf{1}_F$ that takes the values 1 on F and 0 on the complement F^c of F. Events are commonly denoted by brackets surrounding a statement, for instance, if Y is a V-valued random variable, we write $\{Y = v\}$ rather than $\{\omega \in \Omega : Y(\omega) = v\}$ or $Y^{-1}(v)$. The term *almost sure* (abbreviated *a.s.*) is used in place of *almost everywhere*: thus, for instance, we say that $X = c$ almost surely if the event $\{X \neq c\}$ has probability 0.

The integral of a real-valued random variable Y with respect to the measure P, when it exists, is denoted by EY (or sometimes $E_P Y$, if there is a need to specify the probability measure with respect to which the integral refers), and called the *expectation* of Y: thus,

$$EY = \int Y \, dP = \int Y(\omega) \, dP(\omega). \tag{A.1.1}$$

Since the expectation operator is just an instance of the Lebesgue integral, it inherits all of the properites of the integral: *linearity, monotonicity, monotone and dominated convergence theorems,* etc. A random variable X is said to be *integrable* if $E|X| < \infty$.

Any random variable $Y : \Omega \to \Upsilon$ on a probability space $(\Omega, \mathcal{F}, P)$ taking values in a measurable space $(\Upsilon, \mathcal{G})$ induces a probability measure $P \circ Y^{-1}$ on $\mathcal{G}$ called the

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5

distribution of Y, or sometimes the *induced measure* or *pushforward* of P. When $\mathbf{Y} = (Y_n)_{n\geq 0}$ is a sequence of random variables $Y_n : \Omega \to \Upsilon$, the measure $P \circ \mathbf{Y}^{-1}$ is sometimes called the *joint distribution* or the *law* of the sequence. If two random variables Y, Y' valued in Υ (but not necessarily defined on the same probability space) have the same distribution, then we say that Y' is a *version* of Y. If Y is a random variable valued in a measurable space $(\Upsilon, \mathcal{G})$ then any event $F \in \sigma(Y)$ is of the form $F = Y^{-1}(G)$ for some $G \in \mathcal{G}$, and its probability

$$P(F) = (P \circ Y^{-1})(G)$$

is determined by the distribution $P \circ Y^{-1}$ of Y. This is obvious (in fact, tautological) in the abstract, but in concrete instances sometimes causes confusion for newcomers to probability theory, because in many calculations we move back and forth from one probability space to another via induced measures.

A.2 Sigma Algebras

In the world of probability, σ-algebras play a more prominent role than in some other parts of analysis. This is especially true in the theory of *martingales* (see Chapter 8). You should recall that the intersection of any collection of σ-algebras on a set Ω is a σ-algebra; this makes it possible to specify the *minimal* σ-algebra with certain properties. For instance, if $\mathcal{A}$ is any collection of subsets of Ω then there is a minimal σ-algebra on Ω that contains $\mathcal{A}$; this is denoted by

$$\sigma(\mathcal{A}) = \text{minimal } \sigma - \text{algebra containing every } A \in \mathcal{A}. \tag{A.2.1}$$

Similarly, if $(Y_\lambda)_{\lambda\in\Lambda}$ is a collection of random variables defined on $(\Omega, \mathcal{F})$, then we define the *σ-algebra generated by* $(Y_\lambda)_{\lambda\in\Lambda}$,denoted by $\sigma((Y_\lambda)_{\lambda\in\Lambda})$, to be the minimal σ-algebra $\mathcal{G}$ such that each Y_λ is $\mathcal{G}$-measurable. This will, of course, be contained in the reference σ-algebra $\mathcal{F}$. For a single random variable Y taking values in a measurable space $(\Upsilon, \mathcal{G})$ the σ-algebra $\sigma(Y)$ generated by Y is just $Y^{-1}(\mathcal{G})$.

It is sometimes necessary to prove that every element of a σ-algebra $\mathcal{F}$ has a certain property: for instance, to show that two probability measures P, Q on $\mathcal{F}$ are identical, one must prove that $P(F) = Q(F)$ for every $F \in \mathcal{F}$. Such chores are often facilitated by the use of the *Monotone Class Theorem*, which reduces the problem to showing that (i) the property holds for every element in an *algebra* $\mathcal{A}$ that generates $\mathcal{F}$, and (ii) that the collection of events for which the property holds is a monotone class. Let's recall the definitions.

Definition A.2.1 A collection $\mathcal{M}$ of subsets of Ω is a *monotone class* if it is closed under countable monotone unions and intersections, that is, if for every countable subset $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{M}$

$$A_1 \subset A_2 \subset \cdots \quad \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{M} \quad \text{and} \tag{A.2.2}$$

$$A_1 \supset A_2 \supset \cdots \quad \Longrightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{M}. \tag{A.2.3}$$

A collection $\mathcal{A}$ of subsets of Ω is an *algebra* if $\Omega \in \mathcal{A}$ and $\mathcal{A}$ is closed under complements and finite unions, that is, if $A_1, A_2 \in \mathcal{A}$ then $A_1^c \in \mathcal{A}$ and $A_1 \cup A_2 \in \mathcal{A}$.

Proposition A.2.2 (Monotone Class Theorem) *Any monotone class that contains an algebra $\mathcal{A}$ also contains the σ-algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$.*

See, for instance, [121], Lemma 1.7.14 for the somewhat tedious proof.

Corollary A.2.3 *If the σ-algebra $\mathcal{F}$ is generated by an algebra $\mathcal{A}$, then any probability measure on $\mathcal{F}$ is uniquely determined by its values on $\mathcal{A}$.*

Proof. Let P, Q be two probability measures on $\mathcal{F}$ such that $P(A) = Q(A)$ for every event $A \in \mathcal{A}$. Define $\mathcal{M}$ to be the collection of all events $F \in \mathcal{F}$ such that $P(F) = Q(F)$; then by the Dominated Convergence Theorem, $\mathcal{M}$ is a monotone class. Therefore, by the Monotone Class Theorem, $\mathcal{M} = \mathcal{F}$. □

Corollary A.2.4 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $\mathcal{A}$ is an algebra that generates $\mathcal{F}$, then any event $F \in \mathcal{F}$ can be arbitrarily well-approximated by events in A, that is, for any $\varepsilon > 0$ there exists an event $A \in \mathcal{A}$ such that*

$$E|\mathbf{1}_F - \mathbf{1}_A| < \varepsilon. \tag{A.2.4}$$

Proof. Exercise. □

Example A.2.5 Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{A}$ an algebra that generates $\mathcal{F}$. If $X, Y \in L^1(\Omega, \mathcal{F}, P)$ are integrable, real-valued random variables such that

$$EX\mathbf{1}_A = EY\mathbf{1}_A \quad \text{for every } A \in \mathcal{A} \tag{A.2.5}$$

then $X = Y$ almost surely. To see this, consider first the special case where X and Y are both nonnegative. In this case, the Dominated Convergence Theorem implies that the set of events $A \in \mathcal{F}$ for which the equality (A.2.5) holds is a monotone class, so Proposition A.2.2 implies that (A.2.5) holds for all events $A \in \mathcal{F}$. But this implies that for any $n \in \mathbb{N}$ the equality (A.2.5) holds for the events $A = \left\{X > Y + \frac{1}{n}\right\}$ and $A = \left\{X < Y - \frac{1}{n}\right\}$, so these events must have probability 0. It follows that $P\{X > Y\} = P\{Y > X\} = 0$. The general case, where X and Y are not necessarily nonnegative, can be proved by decomposing X and Y into their positive and negative parts and then applying a similar argument. Exercise: Check this.

A.3 Independence

Definition A.3.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space. The events in a subset $\mathcal{S} \subset \mathcal{F}$ are *independent* if for any finite subset $\{F_i\}_{i \in [n]} \subset \mathcal{S}$,

$$P(F_1 \cap F_2 \cap \cdots \cap F_n) = \prod_{i=1}^{n} P(F_i). \tag{A.3.1}$$

Sigma algebras $\{\mathcal{F}_\lambda\}_{\lambda \in \Lambda}$ all contained in $\mathcal{F}$ are said to be independent if for any finite subcollection $\{\mathcal{F}_i\}_{i \in [n]}$ and any selections $F_i \in \mathcal{F}_i$ the events $F_1, F_2, \cdots, F_n$ are independent. Random variables $X_1, X_2, \cdots$ all defined on $(\Omega, \mathcal{F}, P)$, with values in (arbitrary) measurable spaces $(\Upsilon_i, \mathcal{G}_i)$, are independent if the σ-algebras $\sigma(X_1), \sigma(X_2), \cdots$ are independent, that is, if for any choices $G_i \in \mathcal{G}_i$ the events $\{X_i \in G_i\}$ are independent:

$$P\left(\bigcap_{i=1}^{n} \{X_i \in G_i\}\right) = \prod_{i=1}^{n} P\{X_i \in G_i\}. \tag{A.3.2}$$

The condition (A.3.2) can be reformulated in terms of the distributions μ_i of the random variables X_i as follows: the random variables $X_1, X_2, \cdots$ are independent if and only if the (joint) distribution of the sequence $(X_1, X_2, \cdots)$ is the *product measure* $\mu := \mu_1 \times \mu_2 \times \cdots$. This is defined as follows.

Definition A.3.2 Let $(\Upsilon_i, \mathcal{G}_i, \mu_i)_{i \geq 1}$ be a sequence of probability spaces. The *product σ-algebra* $\mathcal{G}^\infty$ on the infinite product space $\Upsilon = \prod_{i \geq 1} \Upsilon_i$ is the smallest σ-algebra containing all *cylinder sets*

$$\prod_{i=1}^{n} G_i \quad \text{where} \quad G_i \in \mathcal{G}_i \quad \text{and} \quad n \in \mathbb{N}. \tag{A.3.3}$$

The *product measure* $\mu = \prod_{i \geq 1} \mu_i$ is the unique probability measure on $\mathcal{G}^\infty$ such that for every cylinder set,

$$\mu\left(\prod_{i=1}^{n} G_i\right) = \prod_{i=1}^{n} \mu_i(G_i). \tag{A.3.4}$$

The existence and uniqueness of the product measure follows from the *Daniell-Kolmogorov* or *Caratheodory* extension theorems of measure theory. See Royden [111], Chapter 12. In most interesting cases, however, the product measure can be constructed using only the existence of Lebesgue measure on the unit interval, as we will explain in Section A.4.

The defining equation (A.3.4) for a product measure extends to expectations.

Proposition A.3.3 (Product Law) *If $X_1, X_2, \cdots, X_n$ are independent random variables taking values in spaces $(\Upsilon_i, \mathcal{G}_i)$, then for any nonnegative random variables $g_i : \Upsilon_i \to \mathbb{R}$,*

$$E \prod_{i=1}^{n} g_i(X_i) = \prod_{i=1}^{n} E g_i(X_i). \tag{A.3.5}$$

Proof. This is by a standard argument in measure theory, which we will not duplicate elsewhere. First, the identity (A.3.5) holds for indicator functions $g_i = \mathbf{1}_{G_i}$, by the definition (A.3.3) of independence. It therefore holds also for *simple* random variables (finite linear combinations of indicators). Finally, by the Monotone Convergence Theorem, (A.3.5) holds for arbitrary nonnegative functions g_i, because any nonnegative, measurable function g is the monotone limit of the simple functions $g_n := \sum_{k=1}^{4^n} (k/2^n) \mathbf{1}\{k/2^n \leq g < (k+1)/2^n\}$. □

One other fact is worth noting: *independence is preserved by amalgamation*. Here is a precise statement.

Proposition A.3.4 *Let $\{\mathcal{F}_\lambda\}_{\lambda \in \Lambda}$ be independent σ-algebras in a probability space $(\Omega, \mathcal{F}, P)$, let $\Lambda = \cup_{i \in I} \Lambda_i$ be a partition of the index set Λ, and for each $i \in I$ let $\mathcal{G}_i = \sigma(\cup_{\lambda \in \Lambda_i} \mathcal{F}_\lambda)$ be the smallest σ-algebra containing all of the σ-algebras $\mathcal{F}_\lambda$ in the ith slice Λ_i of the partition. Then the σ-algebras $\{\mathcal{G}_i\}_{i \in I}$ are independent.*

Proof. This is a routine exercise in the use of the Monotone Class Theorem. This implies that it suffices to show that for any finite collection of events $F_{i,k}$ such that $F_{i,k} \in \mathcal{F}_{\lambda(i,k)}$, where the indices $\lambda(i,k)$ are distinct elements of Λ_i,

$$P\left(\bigcap_{i,k} F_{i,k}\right) = \prod_i P\left(\bigcap_k F_{i,k}\right).$$

But this follows by the independence of the σ-algebras $\mathcal{F}_\lambda$. □

Thus, for example, if $X_1, X_2, \cdots$ are independent binary random variables, each with the Bernoulli$(1/2)$ distribution, then the random variables

$$U_1 := \sum_{n=1}^{\infty} X_{2n}/2^n \quad \text{and}$$

$$U_2 := \sum_{n=1}^{\infty} X_{2n+1}/2^n \tag{A.3.6}$$

are independent uniform$-[0, 1]$ random variables, that is random variables each with the *uniform distribution* (i.e., Lebesgue measure) on the unit interval $[0, 1]$.

A.4 Lebesgue Space

It is important to understand that virtually all interesting random processes — including random walks on countable groups, and more generally Markov chains on countable state spaces — can be constructed on any probability space $(\Omega, \mathcal{F}, P)$ with an infinite sequence $(U_n)_{n\geq 1}$ of independent uniform$-[0, 1]$ random variables. Suppose, for instance, that you want an infinite sequence $\xi_1, \xi_2, \cdots$ of independent random variables taking values in the set $\{1, 2, 3\}$ with probabilities p_a, p_b, p_c, respectively; this can be arranged by partitioning the interval $[0, 1]$ into three nonoverlapping intervals J_1, J_2, J_3 of lengths p_a, p_b, p_c and then setting

$$\xi_n = \sum_{k=1}^{3} k\mathbf{1}_{J_k}(U_n). \tag{A.4.1}$$

The resulting random sequence $\boldsymbol{\xi} := (\xi_n)_{n\geq 1}$ maps the probability space Ω to the product space $\{1, 2, 3\}^\infty$, and this mapping is measurable relative to the Borel σ-algebra $\mathcal{B}$ on $\{1, 2, 3\}^\infty$. The distribution of $\boldsymbol{\xi}$ — i.e., the induced measure $\nu := P \circ \boldsymbol{\xi}^{-1}$ — is the product measure on $(\{1, 2, 3\}^\infty, \mathcal{B})$.

Lebesgue space (the probability space $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$, where λ is Lebesgue measure) supports an infinite sequence of independent uniform$-[0, 1]$ random variables. To see this, observe first that U =id is uniform$-[0, 1]$ (duh!), and then that the binary coefficients $Y_1, Y_2, \cdots$ of $U = \sum_{n\geq 1} 2^{-n} Y_n$ are independent Bernoulli $1/2$ random variables. The infinite sequence $(Y_n)_{n\geq 1}$ can be re-assembled to give an infinite double array $(Y_{m,n})_{m,n\geq 1}$ of independent Bernoullis, and these then used to construct independent uniform$-[0, 1]$ r.v.s, as in (A.3.6). (The amalgamation principle guarantees that the resulting uniforms are actually independent.) Together with the construction of the preceding paragraph, these observations show that the existence of product measures follows directly from the existence of Lebesgue measure; there is no need to invoke the Caratheodory or Daniell-Kolmogorov extension theorems. Similar constructions can be used for Markov chains.

A.5 The Borel-Cantelli Lemma

Dealing with infinite random walk trajectories often requires that we be able to estimate probabilities of events involving infinite unions and intersections of simpler events. One such class of events deserves special mention, as it arises frequently. If $F_1, F_2, \cdots$ is an infinite sequence of events, define the event

$$\{F_n \text{ i.o.}\} = \{F_n \text{ infinitely often}\} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} F_n.$$

Borel-Cantelli Lemma *If* $\sum_{n=1}^{\infty} P(F_n) < \infty$ *then* $P\{F_n \text{ i.o.}\} = 0$.

Proof. Expectation commutes with infinite sums when the terms are all nonnegative, so $\sum_{n=1}^{\infty} P(F_n) = E \sum_{n=1}^{\infty} \mathbf{1}_{F_n} := EN$, where N is the number of events in the sequence $(F_n)_{n\geq 1}$ that occur. If $EN < \infty$ then $P\{N < \infty\} = 1$. □

Corollary A.5.1 *Let $Y_1, Y_2, \cdots$ be identically distributed random variables with finite first moment $E|Y_1| < \infty$. Then with probability* 1,

$$\lim_{n\to\infty} \frac{Y_n}{n} = 0. \tag{A.5.1}$$

Proof. Let's begin by taking note of the following elementary formula for the expectation of a nonnegative, integer-valued random variable Y:

$$EY = E \sum_{n=1}^{\infty} \mathbf{1}\{Y \geq n\} = \sum_{n=1}^{\infty} P\{Y \geq n\}. \tag{A.5.2}$$

Next, fix a scalar $\varepsilon > 0$, and for each n set $Z_n = \lceil |Y_n|/\varepsilon \rceil$. Since the random variables Y_n are identically distributed, so are the random variables Z_n; and since the random variables Y_n have finite first moment, so do the random variables Z_n. Hence,

$$\infty > EZ_1 = \sum_{n=1}^{\infty} P\{Z_1 \geq n\} = \sum_{n=1}^{\infty} P\{Z_n \geq n\} \geq \sum_{n=1}^{\infty} P\{|Y_n| \geq n\varepsilon\}.$$

It now follows by the Borel-Cantelli Lemma that with probability one only finitely many of the events $\{|Y_n| \geq n\varepsilon\}$ will occur, and so

$$P\left\{\limsup_{n\to\infty} |Y_n|/n > \varepsilon\right\} = 0.$$

Since this holds for every εin the countable set $\{m^{-1} : m \in \mathbb{N}\}$, the exceptional sets of probability 0 add up to a set of measure zero, and so

$$P\left\{\limsup_{n\to\infty} |Y_n|/n \leq m^{-1} \text{ for every } m \in \mathbb{N}\right\} = 0.$$

This implies (A.5.1). □

A.6 Hoeffding's Inequality

Proposition A.6.1 (Hoeffding) *Let $Y_1, Y_2, \cdots$ be independent real-valued random variables such that $EY_n = 0$ and $|Y_n| \leq 1$. Let $S_n = \sum_{i \leq n} Y_i$ be the nth partial sum. Then for any $\alpha > 0$ and every $n = 1, 2, \cdots$,*

$$P\{|S_n| \geq \alpha\} \leq 2\exp\left\{-\alpha^2/2n\right\} \quad \Longleftrightarrow \tag{A.6.1}$$

$$P\{|S_n| \geq n\alpha\} \leq 2\exp\left\{-n\alpha^2/2\right\} \tag{A.6.2}$$

Proof. The function $y \mapsto e^y$ is convex, so for any $y \in [-1, 1]$ and any $\theta \in \mathbb{R}$,

$$e^{\theta y} \leq \frac{1}{2}(1+y)e^{\theta} + \frac{1}{2}(1-y)e^{-\theta} = \cosh\theta + \frac{1}{2}(ye^{\theta} - ye^{-\theta}).$$

Substituting Y_i for y and using the hypothesis that $EY_i = 0$, we obtain the inequality $Ee^{\theta Y_i} \leq \cosh\theta$. Consequently, by the product rule for expectations,

$$E\exp\{\theta S_n\} = \prod_{i=1}^{n} E\exp\{\theta Y_i\} \leq (\cosh\theta)^n.$$

Now for any $\theta > 0$ the indicator $\mathbf{1}\{S_n \geq \alpha\}$ is pointwise dominated by the random variable $e^{\theta S_n}/e^{\theta\alpha}$, and similarly the indicator $\mathbf{1}\{S_n \leq -\alpha\}$ is dominated by $e^{-\theta S_n}/e^{\theta\alpha}$. Thus, by the monotonicity property of expectation,

$$P\{S_n \geq \alpha\} \leq \frac{E\exp\{\theta S_n\}}{e^{\theta\alpha}} \leq e^{-\theta\alpha}(\cosh\theta)^n \quad \text{and}$$

$$P\{S_n \leq -\alpha\} \leq \frac{E\exp\{-\theta S_n\}}{e^{\theta\alpha}} \leq e^{-\theta\alpha}(\cosh\theta)^n.$$

Since $\cosh\theta \leq \exp\left\{\theta^2/2\right\}$ (proof: compare Taylor series), it follows that for every $\theta > 0$

$$P\{|S_n| \geq \alpha\} \leq 2\exp\{-\theta\alpha\}\exp\left\{n\theta^2/2\right\}.$$

Setting $\theta = \alpha/n$, we obtain the inequality (A.6.1). □

Proof of SLLN for Bounded r.v.s. The second inequality (A.6.2) shows that the probability of a "large deviation" $|S_n| \geq n\alpha$ decays exponentially in n. In particular, if $Y_1, Y_2, \cdots$ are independent, identically distributed random variables with mean $EY_i = 0$ that satisfy $|Y_i| \leq 1$ then for any $\varepsilon > 0$.

$$\sum_{n=1}^{\infty} P\{|S_n| \geq n\varepsilon\} \leq \sum_{n=1}^{\infty} 2e^{-n\varepsilon^2/2} < \infty.$$

Consequently, by the Borel-Cantelli lemma, the event $\{|S_n| \geq n\varepsilon \text{ i.o.}\}$ has probability 0. Since this is true for every (rational) $\varepsilon > 0$, it follows that with probability one the limsup of the sequence $|S_n|/n$ is 0. This proves the Strong Law of Large Numbers in the special case where the summands Y_i are bounded. □

A.7 Weak Convergence

The notion of *weak convergence* is fundamental to Furstenberg's theory of Poisson boundaries (Chapter 12). Following is a resumé of some of the basic facts that will be needed.

Definition A.7.1 The *Borel σ-algebra* $\mathcal{B} = \mathcal{B}_{\mathcal{Y}}$ on a metric space $\mathcal{Y}$ is the smallest σ-algebra that contains all the open sets; a *Borel measure* is a measure on the σ-algebra $\mathcal{B}$. The set of all Borel probability measures on a metric space $\mathcal{Y}$ is denoted by $\mathcal{M}(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$.

Definition A.7.2 A sequence $(\lambda_n)_{n\geq 1}$ of finite Borel measures on a metric space $(\mathcal{Y}, d)$ *converges weakly* to a finite Borel measure λ if and only if for every bounded, continuous function $f : \mathcal{Y} \to \mathbb{R}$,

$$\lim_{n\to\infty} \int f\, d\lambda_n = \int f\, d\lambda \tag{A.7.1}$$

(Every continuous function on $\mathcal{Y}$ is Borel measurable, so the integrals in (A.7.1) are well-defined.) A sequence $(Y_n)_{n\in\mathbb{N}}$ of $\mathcal{Y}$-valued random variables *converges in distribution* to λ if their distributions converge weakly to λ.

The most important instance of weak convergence is the weak convergence of renormalized sums to the normal (Gaussian) distribution.

Theorem A.7.3 (Central Limit Theorem) *Let $Y_1, Y_2, \cdots$ be independent, identically distributed real-valued random variables with expectation $EY_i = 0$ and positive, finite variance $\sigma^2 := EY_i^2$. Set $S_n = \sum_{i=1}^n Y_i$; then the random variables $S_n/\sqrt{n}$ converge in distribution to the normal distribution λ_σ with mean 0 and variance σ^2, that is, the Borel probability measure*

$$\lambda_\sigma(B) = \frac{1}{\sqrt{2\pi}\sigma} \int_B e^{-x^2/2\sigma^2}\, dx. \tag{A.7.2}$$

For Bernoulli random variables Y_i, the Central Limit Theorem is a consequence of Stirling's formula: see, for instance, [39]. For a proof of a more general theorem — Lindeberg's Central Limit Theorem — see [83], Chapter 6.

Working with weak convergence requires some additional information about the space $C(\mathcal{Y})$ of continuous, real-valued functions on a compact metric space $\mathcal{Y}$.

Theorem A.7.4 (Riesz-Markov) *Let $\mathcal{Y}$ be a compact metric space. For every positive, linear functional $\Lambda : C(\mathcal{Y}) \to \mathbb{R}$ such that $\Lambda(1) = 1$ there exists a unique Borel probability measure λ on $\mathcal{Y}$ such that*

$$\Lambda(f) = E_\lambda f = \int_{\mathcal{Y}} f(y)\, d\lambda(y) \tag{A.7.3}$$

for each $f \in C(\mathcal{Y})$.

See [111], Chapter 13, Theorem 23.

Theorem A.7.5 (Stone-Weierstrass) *Let $\mathcal{Y}$ be a compact metric space and let $A \subset C(\mathcal{Y})$ be an* algebra *of continuous functions over the rational field $\mathbb{Q}$ that contains the constant function $f \equiv 1$. If A* separates points *of $\mathcal{Y}$, then A is dense in $C(\mathcal{Y})$, that is, for any $g \in C(\mathcal{Y})$ and any real $\varepsilon > 0$ there exists $f \in A$ such that*

$$\|f - g\|_\infty := \max_{y \in \mathcal{Y}} |f(y) - g(y)| < \varepsilon. \tag{A.7.4}$$

Note: Here $\|\cdot\|_\infty$is the sup norm, defined by $\|f\|_\infty := \sup_{x \in \mathcal{Y}} |f(x)|$. An *algebra* over a field is a vector space over the field that is closed under multiplication. A collection $B \subset C(\mathcal{Y})$ *separates points* of $\mathcal{Y}$ if for any two distinct points $y, y' \in \mathcal{Y}$ there exists a function $f \in B$ such that $f(y) \neq f(y')$. The Stone-Weierstrass theorem is usually stated and proved for algebras over the field $\mathbb{R}$; see, for instance, Royden [111], Chapter 9. The standard version implies the version for $\mathbb{Q}$-algebras stated above, because the closure of a $\mathbb{Q}$-algebra A in $C(\mathcal{Y})$ is an algebra over $\mathbb{R}$ that contains all real constant functions.

Corollary A.7.6 *If $\mathcal{Y}$ is a compact metric space then the space $C(\mathcal{Y})$ is separable, that is, there is a countable subset $A \subset C(\mathcal{Y})$ that is dense in the sup-norm metric.*

Proof. Any compact metric space is separable, so $\mathcal{Y}$ has a countable dense subset D. For each rational $q \geq 1$ and each $x \in D$ define $f_{x,q} \in C(\mathcal{Y})$ by

$$f_{x,q}(y) = \max(1 - q\,\mathrm{dist}(x, y), 0) \quad \text{for all} \quad y \in \mathcal{Y},$$

where dist is the metric on $\mathcal{Y}$. The set $\{f_{x,q}\}_{x \in D, q \in \mathbb{Q}_{\geq 1}}$ is countable and, since D is dense in $\mathcal{Y}$, this collection of functions separates points of $\mathcal{Y}$. Now let A be the rational algebra generated by the functions $f_{x,q}$ and the rational constant functions, that is, A is the set of all functions of the form

$$\sum_{j=1}^{n} \prod_{i=1}^{n(j)} (a_{i,j} f_{x_{i(j)}, q_{i(j)}} + b_{i,j})$$

where $a_{i(j)}, b_{i(j)}$ are rational numbers. The set A is countable, and by construction it satisfies the requirements of the Stone-Weierstrass theorem. □

Exercise A.7.7 Let $(\mathcal{Y}, d)$ be a compact metric space, and denote by $C(\mathcal{Y})$ the Banach space of continuous, real-valued functions on $\mathcal{Y}$, with the supremum norm $\|\cdot\|_\infty$. Show that if $D = \{f_n\}_{n\in\mathbb{N}}$ is a dense subset of $C(\mathcal{Y})$, then a sequence λ_n of Borel probability measures on $\mathcal{Y}$ converges weakly to a Borel probability measure λ if and only if for every $f_k \in D$,

$$\lim_{n\to\infty} E_{\lambda_n} f_k = E_\lambda f_k.$$

Corollary A.7.8 *If $\mathcal{Y}$ is a compact metric space then the topology of weak convergence on the space $\mathcal{M}(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ of Borel probability measures is metrizable.*

Proof. Let $D = \{f_n\}_{n\in\mathbb{N}}$ be a dense subset of $C(\mathcal{Y})$ that does not include the zero function. For any two probability measures $\alpha, \beta \in \mathcal{M}(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ define

$$d(\alpha, \beta) = \sum_{n=1}^{\infty} \frac{1}{2^n \|f_n\|_\infty} \left| \int f_n \, d\alpha - \int f_n \, d\beta \right|;$$

this is evidently symmetric, nonnegative, and satisfies the triangle inequality. If $d(\alpha, \beta) = 0$ then $\int f_n \, d\alpha = \int f_n \, d\beta$ for every $f_n \in D$, and since D is dense in $C\mathcal{Y}$ this implies that

$$\int f \, d\alpha = \int f \, d\beta \quad \text{for all } f \in C(\mathcal{Y}).$$

The Riesz-Markov Theorem implies that $d(\alpha, \beta) = 0$ only if $\alpha = \beta$. Hence, d is a metric on $\mathcal{M}(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$. Exercise A.7.7 shows that weak convergence is equivalent to convergence in the metric d. □

Theorem A.7.9 (Helly-Prohorov) *If the metric space $\mathcal{Y}$ is a compact, metrizable topological space then every sequence λ_n of Borel probability measures on $\mathcal{Y}$ has a weakly convergent subsequence, that is, there exist a subsequence $\lambda_m = \lambda_{n_m}$ and a Borel probability measure λ such that* (A.7.1) *holds for every continuous function $f : \mathcal{Y} \to \mathbb{R}$. Consequently, the topology of weak convergence on $\mathcal{M}(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ is compact.*

Proof. By Corollary A.7.6 the space $C(\mathcal{Y})$ has a countable dense subset $D = \{f_k\}_{k\in\mathbb{N}}$, which without loss of generality we can take to be a vector space over the rationals $\mathbb{Q}$. For any element $f_k \in D$, the sequence $E_{\lambda_n} f_k$ of expectations

is bounded in absolute value (by $\|f_k\|_\infty$), so the Bolzano-Weierstrass Theorem implies that any subsequence of $(\lambda_n)_{n\in\mathbb{N}}$ has a subsequence $(\lambda_m)_{m\in\mathbb{N}}$ such that $\lim_{m\to\infty} E_{\lambda_m} f_k$ exists. Hence, by Cantor's diagonal trick, there is a subsequence $(\lambda_m)_{m\in\mathbb{N}}$ such that

$$\Lambda(f_k) := \lim_{m\to\infty} E_{\lambda_m} f_k$$

exists for every $f_k \in D$. The functional Λ is linear and positive on D, and satisfies $\Lambda(1) = 1$. Since D is dense in $C(\mathcal{Y})$, Λ extends to a positive, linear functional on all of $C(\mathcal{Y})$. Therefore, by the Riesz-Markov Theorem, there exists a Borel probability measure λ such that

$$\Lambda(f) = \int f\, d\lambda \quad \text{for every} \quad f \in C(\mathcal{Y}).$$

Weak convergence of the subsequence to the probability measure λ now follows by Exercise A.7.7. □

A.8 Radon-Nikodym Theorem and Likelihood Ratios

Definition A.8.1 Let μ and λ be finite measures on a common measurable space $(\Omega, \mathcal{F})$. The measure μ is *absolutely continuous* with respect to λ on the σ-algebra $\mathcal{F}$, denoted $\mu \ll \lambda$, if every λ-null event is a μ-null event, i.e., for any $F \in \mathcal{F}$,

$$\lambda(F) = 0 \quad \Longrightarrow \quad \mu(F) = 0. \tag{A.8.1}$$

The probability measures λ and μ are *singular* (denoted $\mu \perp \lambda$) if there is an event $F \in \mathcal{F}$ such that

$$\lambda(F^c) = 0 \quad \text{and} \quad \mu(F) = 0. \tag{A.8.2}$$

Example A.8.2 Let X be a nonnegative random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ such that $EX = 1$. For any $F \in \mathcal{F}$, define

$$Q(F) = E(X\mathbf{1}_F). \tag{A.8.3}$$

Then Q is absolutely continuous with respect to P on $\mathcal{F}$.

Example A.8.3 Absolute continuity and singularity of probability measures depend on the reference σ-algebra: a probability measure Q can be singular to P on $\mathcal{F}$ but nevertheless absolutely continuous with respect to P on a smaller σ-algebra $\mathcal{G} \subset \mathcal{F}$. Here is an example: let $\Omega = \{0, 1\}^{\mathbb{N}}$ be the space of all infinite binary sequences, and let $\mathcal{F}$ be the Borel σ-algebra on Ω, that is, $\mathcal{F} = \sigma(\xi_1, \xi_2, \cdots)$

where $\xi_i : \Omega \to \{0, 1\}$ is the ith coordinate evaluation map. For each $n \in \mathbb{N}$ let $\mathcal{F}_n = \sigma(\xi_1, \xi_2, \cdots, \xi_n)$; thus,

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}.$$

By the Daniell-Kolmogorov extension theorem, for each real number $p \in (0, 1)$ there is a unique probability measure P_p on $\mathcal{F}$, the *product-Bernoulli-p* measure, such that for every cylinder event,

$$P_p(C(x_1, x_2, \cdots, x_n)) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}. \tag{A.8.4}$$

These measures are mutually singular on $\mathcal{F}$, but any two are mutually absolutely continuous on every $\mathcal{F}_n$.

Exercise A.8.4 Explain why.

HINT: The Strong Law of Large Numbers might be relevant.

Exercise A.8.5 Show that Q is absolutely continuous with respect to P on $\mathcal{F}$ if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all events $F \in \mathcal{F}$,

$$P(F) < \delta \quad \Longrightarrow \quad Q(F) < \varepsilon. \tag{A.8.5}$$

HINT: If (A.8.5) fails, there exist $\varepsilon > 0$ and events $F_n \in \mathcal{F}$ such that $P(F_n) < 2^{-n}$ but $Q(F_n) > \varepsilon$. What can you say about

$$G := \bigcap_{n \geq 1} \bigcup_{m \geq n} F_m?$$

Theorem A.8.6 (Radon-Nikodym) *For any two finite nonnegative measures μ, λ on a measurable space $(\Omega, \mathcal{F})$ there is a unique decomposition*

$$\mu = \mu_{ac} + \mu_s \tag{A.8.6}$$

of μ into nonnegative measures μ_{ac}, μ_s such that $\mu_{ac} \ll \lambda$ and $\mu_s \perp \lambda$. If μ is absolutely continuous with respect to λ on the σ-algebra $\mathcal{F}$ then there exists a unique (up to change on sets of λ-measure 0), nonnegative, $\mathcal{F}$-measurable function $L := (d\mu/d\lambda)_{\mathcal{F}}$ such that for every $F \in \mathcal{F}$,

$$\mu(F) = \int L \mathbf{1}_F \, d\lambda \tag{A.8.7}$$

Consequently, for any nonnegative $\mathcal{F}$-measurable function Z,

$$\int Z \, d\mu = \int Z L \, d\lambda \tag{A.8.8}$$

The function L is called Radon-Nikodym derivative*, of μ with respect to λ.*

See or Royden [111], Chapter 11 for a proof, or Rudin [112], Chapter 6 for a different approach. The integral formula (A.8.8) follows directly from (A.8.7) (see the proof of Proposition A.3.3). A routine argument, using the formula (A.8.8), shows that Radon-Nikodym derivatives obey a *multiplication rule*: if $\nu \ll \mu \ll \lambda$, then

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}. \tag{A.8.9}$$

It is important to realize that the Radon-Nikodym derivative $L = d\mu/d\lambda$ is measurable with respect to the σ-algebra in question: if the measures λ, μ are restricted to a smaller σ-algebra $\mathcal{G} \subset \mathcal{F}$ then the Radon-Nikodym derivative will (generally) change.

When $\lambda = P$ and $\mu = Q$ are probability measures, a different terminology is used: the Radon-Nikodym derivative dQ/dP, when $Q \ll P$, is called the *likelihood ratio* of Q with respect to P. In this case, the identity (A.8.8) becomes

$$E_Q Z = E_P(ZL). \tag{A.8.10}$$

Exercise A.8.7 Let μ and λ be finite measures on $(\Omega, \mathcal{F})$, and let $T : \Omega \to \Omega$ be a bijection such that both T and T^{-1} are measurable relative to $\mathcal{F}$. Show that if $\mu \ll \lambda$ then $\mu \circ T^{-1} \ll \lambda \circ T^{-1}$ and

$$\frac{d\mu \circ T^{-1}}{d\lambda \circ T^{-1}}(\omega) = \frac{d\mu}{d\lambda}(T^{-1}(\omega)).$$

Exercise A.8.8 Let μ and λ be finite measures on $(\Omega, \mathcal{F})$, and let $T : \Omega \to \Omega'$ be a measurable transformation with range $(\Omega', \mathcal{F}')$. Show that if the induced measure $\mu \circ T^{-1}$ on $\mathcal{F}'$ is absolutely continuous with respect to $\lambda \circ T^{-1}$, then $\mu \ll \lambda$ on the σ-algebra $T^{-1}(\mathcal{F}')$, and that the Radon-Nikodym derivatives satisfy

$$\frac{d\mu \circ T^{-1}}{d\lambda \circ T^{-1}}(T(\omega)) = \left(\frac{d\mu}{d\lambda}\right)_{T^{-1}(\mathcal{F}')}((\omega)).$$

Also, show by example that μ need not be absolutely continuous with respect to λ on the full σ-algebra $\mathcal{F}$.

A.9 Conditional Expectation

Corollary A.9.1 (Conditional Expectation) *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra contained in $\mathcal{F}$. For any real-valued, $\mathcal{F}$-measurable,*

integrable random variable X, there exists a unique (up to changes on events of probability 0*) $\mathcal{G}$-measurable, integrable random variable Y such that for every event $G \in \mathcal{G}$,*

$$E(X\mathbf{1}_G) = E(Y\mathbf{1}_G). \tag{A.9.1}$$

The mapping $X \mapsto Y$ is linear and monotone, and fixes $\mathcal{G}$-measurable random variables.

Terminology. The random variable Y is called the *conditional expectation* of X with respect to the σ-algebra $\mathcal{G}$, and usually denoted by $Y = E(X \mid \mathcal{G})$. The identity (A.9.1) holds, in particular, with $G = \Omega$, so conditional expectation satisfies the identity

$$EE(X \mid \mathcal{G}) = EX. \tag{A.9.2}$$

Proof. Proof of Corollary A.9.1. It suffices to consider the case where $X \geq 0$, as the general case can then be obained by decomposing $X = X_+ - X_-$into its positive and negative parts. Furthermore, by scaling it can be arranged that $EX = 1$. Now if X is nonnegative and integrates to 1 then by Example A.8.3 it is the likelihood ratio of the probability measure $Q(F) := E(X\mathbf{1}_F)$, which is absolutely continuous with respect to P on $\mathcal{F}$, and hence also on $\mathcal{G}$. Therefore, the Radon-Nikodym theorem provides a random variable $Y = (dQ/dP)_{\mathcal{G}}$ such that equation (A.9.1) holds for all $G \in \mathcal{G}$. The linearity and monotonicity of the mapping $X \mapsto Y$ follow by the uniqueness of Radon-Nikodym derivatives, as does the fact that $\mathcal{G}$-measurable random variables are fixed. □

Example A.9.2 If $\mathcal{G} \subset \mathcal{F}$ is a *finite* σ-algebra whose only event of probability 0 is the empty set $\emptyset$ then it is generated by a finite measurable partition $\gamma = (G_i)_{1 \leq i \leq I}$ of the probability space Ω, consisting of those events $G_i \in \mathcal{G}$ of positive probability with no proper subsets $G \in \mathcal{G}$ other than $\emptyset$. In this case, conditional expectations have the following explicit form: for any integrable random variable X,

$$E(X \mid \mathcal{G}) = \sum_{i=1}^{I} \frac{E(X\mathbf{1}_{G_i})}{P(G_i)} \mathbf{1}_{G_i}. \tag{A.9.3}$$

Exercise A.9.3 Let $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ be σ-algebras in a probability space $(\Omega, \mathcal{F}, P)$ such that $\mathcal{G}_2$ is a *null extension* of $\mathcal{G}_1$, that is, for every event $G \in \mathcal{G}_2$ there is an event $G' \in \mathcal{G}_1$ such that $P(G \Delta G') = 0$. Show that for every integrable real random variable X,

$$E(X \mid \mathcal{G}_1) = E(X \mid \mathcal{G}_2) \quad \text{almost surely.} \tag{A.9.4}$$

Exercise A.9.4 Show that if an integrable random variable X is independent of a σ-algebra $\mathcal{G} \subset \mathcal{F}$ (that is, the σ-algebra $\sigma(X)$ generated by X is independent of $\mathcal{G}$),

then

$$E(X \mid \mathcal{G}) = EX \quad \text{almost surely.} \tag{A.9.5}$$

Exercise A.9.5 Show that if X has finite second moment, then its conditional expectation $E(X \mid \mathcal{G})$ on the σ-algebra $\mathcal{G} \subset \mathcal{F}$ is the L^2-orthogonal projection on the closed subspace $L^2(\Omega, \mathcal{G}, P)$ of $L^2(\Omega, \mathcal{F}, P)$ consisting of those square-integrable random variables that are measurable with respect to $\mathcal{G}$. Conclude that

(a) $E(E(X \mid \mathcal{G})^2) = EX^2 < \infty$ if and only if X is almost surely equal to a $\mathcal{G}$-measurable random variable; and
(b) for any event F, the conditional expectation $E(\mathbf{1}_F \mid \mathcal{G})$ is almost surely equal to an indicator random variable $\mathbf{1}_G$ if and only if $\mathbf{1}_F = \mathbf{1}_G$ almost surely.

Properties of Conditional Expectation. The following are all routine consequences of Corollary A.9.1, as this guarantees not only the existence of a $\mathcal{G}$-measurable random variable Y satisfying the identity (A.9.1), but also its essential uniqueness. In each statement, equality of random variables should be interpreted to mean almost sure equality.

(CE0) $E(XZ) = E(E(X|\mathcal{G})Z)$ for any bounded, $\mathcal{G}$-measurable r.v. Z.
(CE1) Linearity: $E(aX_1 + bX_2 \mid \mathcal{G}) = aE(X_1 \mid \mathcal{G}) + bE(X_2 \mid \mathcal{G})$.
(CE2) Positivity: If $X \geq 0$ then $E(X \mid \mathcal{G}) \geq 0$.
(CE3) Monotonicity: If $X_1 \leq X_2$ then $E(X_1 \mid \mathcal{G}) \leq E(X_2 \mid \mathcal{G})$.
(CE4) Tower Property: If $\mathcal{G}_1 \subset \mathcal{G}_2$ then $E(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1) = E(X \mid \mathcal{G}_1)$.
(CE5) Monotone Convergence: If $0 \leq X_1 \leq X_2 \leq \cdots$ and $\lim_{n\to\infty} X_n = X$ then

$$E(X \mid \mathcal{G}) = \lim_{n\to\infty} E(X_n \mid \mathcal{G}).$$

(CE6) Projection Property: If Z is $\mathcal{G}$-measurable and X, XZ are both integrable then

$$E(XZ \mid \mathcal{G}) = ZE(X \mid \mathcal{G}).$$

The Tower Property is especially noteworthy, as it guarantees that for any integrable random variable X and any filtration $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{G}_2 \subset \cdots$ (see Section 8.3), the sequence $(E(X \mid \mathcal{G}_n))_{n\geq 0}$ is a martingale.

Proposition A.9.6 *Let X, Y be independent random variables taking values in the measurable spaces $(\Upsilon_1, \mathcal{G}_1)$ and $(\Upsilon_2, \mathcal{G}_2)$, respectively, and let $F : \Upsilon_1 \times \Upsilon_2 \to \mathbb{R}_+$ be a nonnegative, jointly measurable function. Then*

$$E(F(X,Y) \mid \sigma(X)) = \Phi(X) \quad \textit{where} \quad \Phi(x) := EF(x,Y) \textit{ for all } x \in \Upsilon_1. \tag{A.9.6}$$

Proof. This is an extension of Exercise A.9.4. If F has the form $F(x, y) = f(x)g(y)$ then the identity (A.9.6) follows directly from Exercise A.9.4. Consequently, by the linearity of expectation and conditional expectation, (A.9.6) holds for all simple functions of the form

$$F(x, y) = \sum_{i=1}^{I} \sum_{j=1}^{J} c_{i,j} \mathbf{1}_{G_i}(x) \mathbf{1}_{H_j}(y).$$

Finally, since every nonnegative, jointly measurable function $F : \Upsilon_1 \times \Upsilon_2 \to \mathbb{R}_+$ is a monotone limit of simple functions, (A.9.6) follows by the Monotone Convergence Property of conditional expectation. □

Theorem A.9.7 (Jensen's Inequality for Conditional Expectation) *Let X be an integrable random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ that takes values in an interval $J \subset \mathbb{R}$, and let $\varphi : J \to \mathbb{R}$ be a convex function such that $\varphi(X)$ is integrable. Then for any σ-algebra $\mathcal{G} \subset \mathcal{F}$,*

$$\varphi(E(X \mid \mathcal{G})) \leq E(\varphi(X) \mid \mathcal{G}) \quad \textit{almost surely.} \tag{A.9.7}$$

Moreover, if φ is strictly convex then X is almost surely constant on the event that equality holds in (A.9.7).

Proof. Jensen's Inequality for ordinary expectations (cf. Royden [111], Section V.5) states that under the hypotheses of the theorem, $\varphi(EX) \leq E(\varphi(X))$. Now for any $G \in \mathcal{G}$ with positive probability $P(G) > 0$, the random variable $\mathbf{1}_G/P(G)$ is the likelihood ratio of a probability measure $Q = Q_G$ with respect to P, and so Jensen's inequality applies to Q: in particular,

$$\varphi(E_Q X) \leq E_Q \varphi(X) \quad \Longleftrightarrow \quad \varphi(E(\mathbf{1}_G X)/P(G)) \leq E(\varphi(X)\mathbf{1}_G)/P(G) \tag{A.9.8}$$

where $E = E_P$ denotes expectation under P. Since X and $\varphi(X)$ are integrable, they have conditional expectations with respect to the σ-algebra $\mathcal{G}$; hence, (A.9.8) implies that for every $G \in \mathcal{G}$,

$$\varphi(E(\mathbf{1}_G E(X \mid \mathcal{G}))/P(G)) \leq E(E(\varphi(X) \mid \mathcal{G})\mathbf{1}_G)/P(G).$$

It now follows that the event $G := \{E(\varphi(X) \mid \mathcal{G}) < \varphi(E(X \mid \mathcal{G}))\}$ must have probability 0, because otherwise the last displayed inequality would lead to a contradiction. If φ is strictly convex then Jensen's inequality $\varphi(EX) \leq E\varphi(X)$ for ordinary expectations is strict unless $X = EX$ almost surely. Hence, if φ is strictly convex then for each event $G \in \mathcal{G}$ the inequality (A.9.8) is strict unless X is constant Q_G-almost surely; in particular, X is almost surely constant on the event $G = \{\varphi(E(X \mid \mathcal{G})) = E(\varphi(X) \mid \mathcal{G})\}$. □

References

1. Lars V. Ahlfors. *Complex analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York, third edition, 1978. An introduction to the theory of analytic functions of one complex variable.
2. Roger C. Alperin. PSL2(Z)=Z2*Z3. *Amer. Math. Monthly*, 100(8):385–386, 1993.
3. Alano Ancona. Positive harmonic functions and hyperbolicity. In *Potential theory—surveys and problems (Prague, 1987)*, volume 1344 of *Lecture Notes in Math.*, pages 1–23. Springer, Berlin, 1988.
4. André Avez. Entropie des groupes de type fini. *C. R. Acad. Sci. Paris Sér. A-B*, 275:A1363–A1366, 1972.
5. André Avez. Théorème de Choquet-Deny pour les groupes à croissance non exponentielle. *C. R. Acad. Sci. Paris Sér. A*, 279:25–28, 1974.
6. Laurent Bartholdi, Serge Cantat, Tullio Ceccherini-Silberstein, and Pierre de la Harpe. Estimates for simple random walks on fundamental groups of surfaces. *Colloq. Math.*, 72(1):173–193, 1997.
7. H. Bass. The degree of polynomial growth of finitely generated nilpotent groups. *Proc. London Math. Soc. (3)*, 25:603–614, 1972.
8. Alan F. Beardon. *The geometry of discrete groups*, volume 91 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1983.
9. Itai Benjamini and Yuval Peres. Tree-indexed random walks on groups and first passage percolation. *Probab. Theory Related Fields*, 98(1):91–112, 1994.
10. Yves Benoist and Jean-François Quint. *Random walks on reductive groups*, volume 62 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Cham, 2016.
11. G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci.*, 17:656–660, 1931.
12. Sébastien Blachère, Peter Haïssinsky, and Pierre Mathieu. Asymptotic entropy and Green speed for random walks on countable groups. *Ann. Probab.*, 36(3):1134–1152, 2008.
13. David Blackwell. On transient Markov processes with a countable number of states and stationary transition probabilities. *Ann. Math. Statist.*, 26:654–658, 1955.
14. David Blackwell and David Freedman. The tail σ-field of a Markov chain and a theorem of Orey. *Ann. Math. Statist.*, 35:1291–1295, 1964.
15. David Blackwell and David Kendall. The martin boundary of Pólya's urn scheme, and an application to stochastic population growth. *J. Appl. Probability*, 1:284–296, 1964.
16. Béla Bollobás. *Modern graph theory*, volume 184 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5

17. Philippe Bougerol and Jean Lacroix. *Products of random matrices with applications to Schrödinger operators*, volume 8 of *Progress in Probability and Statistics*. Birkhäuser Boston, Inc., Boston, MA, 1985.
18. Theodor Bröcker and Tammo tom Dieck. *Representations of compact Lie groups*, volume 98 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. Translated from the German manuscript, Corrected reprint of the 1985 translation.
19. Danny Calegari. The ergodic theory of hyperbolic groups. In *Geometry and topology down under*, volume 597 of *Contemp. Math.*, pages 15–52. Amer. Math. Soc., Providence, RI, 2013.
20. Elisabetta Candellero and Lorenz A. Gilch. Phase transitions for random walk asymptotics on free products of groups. *Random Structures Algorithms*, 40(2):150–181, 2012.
21. James W. Cannon. The combinatorial structure of cocompact discrete hyperbolic groups. *Geom. Dedicata*, 16(2):123–148, 1984.
22. Thomas Keith Carne. A transmutation formula for Markov chains. *Bull. Sci. Math. (2)*, 109(4):399–405, 1985.
23. Donald I. Cartwright. Some examples of random walks on free products of discrete groups. *Ann. Mat. Pura Appl. (4)*, 151:1–15, 1988.
24. Gustave Choquet and Jacques Deny. Sur l'équation de convolution $\mu = \mu * \sigma$. *C. R. Acad. Sci. Paris*, 250:799–801, 1960.
25. K. L. Chung and W. H. J. Fuchs. On the distribution of values of sums of random variables. *Mem. Amer. Math. Soc.*, 6:12, 1951.
26. Tobias H. Colding and William P. Minicozzi, II. Harmonic functions on manifolds. *Ann. of Math. (2)*, 146(3):725–747, 1997.
27. Thierry Coulhon. Ultracontractivity and Nash type inequalities. *J. Funct. Anal.*, 141(2):510–539, 1996.
28. Thierry Coulhon and Laurent Saloff-Coste. Isopérimétrie pour les groupes et les variétés. *Rev. Mat. Iberoamericana*, 9(2):293–314, 1993.
29. R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.*, 100(1):32–74, 1928.
30. Mahlon Marsh Day. Convolutions, means, and spectra. *Illinois J. Math.*, 8:100–111, 1964.
31. Pierre de la Harpe. *Topics in Geometric Group Theory*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2000.
32. Yves Derriennic. Quelques applications du théorème ergodique sous-additif. In *Conference on Random Walks (Kleebach, 1979) (French)*, volume 74 of *Astérisque*, pages 183–201, 4. Soc. Math. France, Paris, 1980.
33. W. Doeblin. Exposé de la théorie générale des chaînes simples constantes de Markoff. *Rev. Math. Union Interbalkan*, 2:77–105, 1938.
34. J. L. Doob. *Stochastic Processes*. John Wiley & Sons, Inc., New York; Chapman & Hall, Limited, London, 1953.
35. Peter G. Doyle and J. Laurie Snell. *Random Walks and Electric Networks*, volume 22 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984.
36. E. B. Dynkin and M. B. Maljutov. Random walk on groups with a finite number of generators. *Dokl. Akad. Nauk SSSR*, 137:1042–1045, 1961.
37. Manfred Einsiedler and Thomas Ward. *Ergodic theory with a view towards number theory*, volume 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London, 2011.
38. Anna Erschler. Poisson-Furstenberg boundary of random walks on wreath products and free metabelian groups. *Comment. Math. Helv.*, 86(1):113–143, 2011.
39. William Feller. *An Introduction to Probability Theory and its Applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.
40. Simion Filip. Notes on the multiplicative ergodic theorem. *Ergodic Theory Dynam. Systems*, 39(5):1153–1189, 2019.
41. Erling Følner. On groups with full Banach mean value. *Math. Scand.*, 3:243–254, 1955.
42. David Freedman. *Markov chains*. Springer-Verlag, New York-Berlin, 1983. Corrected reprint of the 1971 original.

43. Joshua Frisch, Yair Hartman, Omer Tamuz, and Pooya Vahidi Ferdowsi. Choquet-deny groups and the infinite conjugacy class property. *Ann. of Math. (2)*, 190(1):307–320, 2019.
44. H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.
45. Harry Furstenberg. Noncommuting random products. *Trans. Amer. Math. Soc.*, 108:377–428, 1963.
46. Harry Furstenberg. A Poisson formula for semi-simple Lie groups. *Ann. of Math. (2)*, 77:335–386, 1963.
47. Harry Furstenberg. Random walks and discrete subgroups of Lie groups. In *Advances in Probability and Related Topics, Vol. 1*, pages 1–63. Dekker, New York, 1971.
48. Harry Furstenberg. Boundary theory and stochastic processes on homogeneous spaces. In *Harmonic analysis on homogeneous spaces (Proc. Sympos. Pure Math., Vol. XXVI, Williams Coll., Williamstown, Mass., 1972)*, pages 193–229, 1973.
49. Peter Gerl and Wolfgang Woess. Local limits and harmonic functions for nonisotropic random walks on free groups. *Probab. Theory Relat. Fields*, 71(3):341–355, 1986.
50. Étienne Ghys and Pierre de la Harpe. Espaces métriques hyperboliques. In *Sur les groupes hyperboliques d'après Mikhael Gromov (Bern, 1988)*, volume 83 of *Progr. Math.*, pages 27–45. Birkhäuser Boston, Boston, MA, 1990.
51. L. A. Gilch. Acceleration of lamplighter random walks. *Markov Process. Related Fields*, 14(4):465–486, 2008.
52. Alexander Gorodnik and Amos Nevo. *The ergodic theory of lattice subgroups*, volume 172 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2010.
53. Alexander Gorodnik and Amos Nevo. Quantitative ergodic theorems and their number-theoretic applications. *Bull. Amer. Math. Soc. (N.S.)*, 52(1):65–113, 2015.
54. Sébastien Gouëzel. Local limit theorem for symmetric random walks in Gromov-hyperbolic groups. *J. Amer. Math. Soc.*, 27(3):893–928, 2014.
55. Sébastien Gouezel. A numerical lower bound for the spectral radius of random walks on surface groups. *Combin. Probab. Comput.*, 24(6):838–856, 2015.
56. Sébastien Gouëzel and Steven P. Lalley. Random walks on co-compact Fuchsian groups. *Ann. Sci. Éc. Norm. Supér. (4)*, 46(1):129–173 (2013), 2013.
57. Sébastien Gouëzel, Frédéric Mathéus, and François Maucourant. Entropy and drift in word hyperbolic groups. *Invent. Math.*, 211(3):1201–1255, 2018.
58. David Griffeath. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:95–106, 1974/75.
59. M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.*, pages 75–263. Springer, New York, 1987.
60. Mikhael Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, (53):53–73, 1981.
61. Y. Guivarc'h. Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire. In *Conference on Random Walks (Kleebach, 1979) (French)*, volume 74 of *Astérisque*, pages 47–98, 3. Soc. Math. France, Paris, 1980.
62. Yves Guivarc'h. Croissance polynomiale et périodes des fonctions harmoniques. *Bull. Soc. Math. France*, 101:333–379, 1973.
63. Peter Haïssinsky, Pierre Mathieu, and Sebastian Müller. Renewal theory for random walks on surface groups. *Ergodic Theory Dynam. Systems*, 38(1):155–179, 2018.
64. P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980.
65. I. N. Herstein. *Topics in Algebra*. Xerox College Publishing, Lexington, Mass.-Toronto, Ont., second edition, 1975.
66. John E. Hopcroft and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation*. Addison-Wesley Series in Computer Science. Addison-Wesley Publishing Co., Reading, Mass., 1979.

67. Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.
68. V. A. Kaimanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.
69. Vadim A. Kaimanovich. The Poisson boundary of hyperbolic groups. *C. R. Acad. Sci. Paris Sér. I Math.*, 318(1):59–64, 1994.
70. Vadim A. Kaimanovich. The Poisson formula for groups with hyperbolic properties. *Ann. of Math. (2)*, 152(3):659–692, 2000.
71. Shizuo Kakutani. Two-dimensional Brownian motion and harmonic functions. *Proc. Imp. Acad. Tokyo*, 20:706–714, 1944.
72. Anders Karlsson and François Ledrappier. On laws of large numbers for random walks. *Ann. Probab.*, 34(5):1693–1706, 2006.
73. Anders Karlsson and François Ledrappier. Linear drift and Poisson boundary for random walks. *Pure Appl. Math. Q.*, 3(4, Special Issue: In honor of Grigory Margulis. Part 1):1027–1036, 2007.
74. Anders Karlsson and François Ledrappier. Noncommutative ergodic theorems. In *Geometry, rigidity, and group actions*, Chicago Lectures in Math., pages 396–418. Univ. Chicago Press, Chicago, IL, 2011.
75. Anders Karlsson and Gregory A. Margulis. A multiplicative ergodic theorem and nonpositively curved spaces. *Comm. Math. Phys.*, 208(1):107–123, 1999.
76. Harry Kesten. Full Banach mean values on countable groups. *Math. Scand.*, 7:146–156, 1959.
77. Harry Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.
78. J. F. C. Kingman. The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. Ser. B*, 30:499–510, 1968.
79. J. F. C. Kingman. Subadditive ergodic theory. *Ann. Probability*, 1:883–909, 1973.
80. Bruce Kleiner. A new proof of Gromov's theorem on groups of polynomial growth. *J. Amer. Math. Soc.*, 23(3):815–829, 2010.
81. Ulrich Krengel. *Ergodic theorems*, volume 6 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1985. With a supplement by Antoine Brunel.
82. Nicolas Kryloff and Nicolas Bogoliouboff. La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. *Ann. of Math. (2)*, 38(1):65–113, 1937.
83. Steven P. Lalley. Lecture notes on measure-theoretic probability. http://galton.uchicago.edu/~lalley/Courses/381/Convolutions-Smoothing.pdf.
84. Steven P. Lalley. Finite range random walk on free groups and homogeneous trees. *Ann. Probab.*, 21(4):2087–2130, 1993.
85. Steven P. Lalley. Random walks on regular languages and algebraic systems of generating functions. In *Algebraic methods in statistics and probability (Notre Dame, IN, 2000)*, volume 287 of *Contemp. Math.*, pages 201–230. Amer. Math. Soc., Providence, RI, 2001.
86. Gregory F. Lawler. Cut times for simple random walk. *Electron. J. Probab.*, 1:no. 13, approx. 24 pp. 1996.
87. Joseph Lehner. *Discontinuous Groups and Automorphic Functions*. Mathematical Surveys, No. VIII. American Mathematical Society, Providence, R.I., 1964.
88. Peter Li. *Geometric Analysis*, volume 134 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2012.
89. Peter Li and Luen-Fai Tam. Linear growth harmonic functions on a complete manifold. *J. Differential Geom.*, 29(2):421–425, 1989.
90. Elon Lindenstrauss. Pointwise theorems for amenable groups. *Invent. Math.*, 146(2):259–295, 2001.
91. Russell Lyons and Yuval Peres. *Probability on Trees and Networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016.

92. Russell Lyons and Yuval Peres. Poisson boundaries of lamplighter groups: proof of the Kaimanovich-Vershik conjecture. *J. Eur. Math. Soc. (JEMS)*, 23(4):1133–1160, 2021.
93. Terry Lyons. A simple criterion for transience of a reversible Markov chain. *Ann. Probab.*, 11(2):393–402, 1983.
94. Joseph Maher and Giulio Tiozzo. Random walks on weakly hyperbolic groups. *J. Reine Angew. Math.*, 742:187–239, 2018.
95. Jean Mairesse and Frédéric Mathéus. Random walks on free products of cyclic groups. *J. Lond. Math. Soc. (2)*, 75(1):47–66, 2007.
96. Pierre Mathieu. Carne-Varopoulos bounds for centered random walks. *Ann. Probab.*, 34(3):987–1011, 2006.
97. John Milnor. Growth of finitely generated solvable groups. *J. Differential Geometry*, 2:447–449, 1968.
98. Bojan Mohar and Wolfgang Woess. A survey on spectra of infinite graphs. *Bull. London Math. Soc.*, 21(3):209–234, 1989.
99. B. Morris and Yuval Peres. Evolving sets, mixing and heat kernel bounds. *Probab. Theory Related Fields*, 133(2):245–266, 2005.
100. Peter Mörters and Yuval Peres. *Brownian motion*, volume 30 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.
101. James R. Munkres. *Topology*. Prentice Hall, Inc., Upper Saddle River, NJ, 2000. Second edition of [MR0464128].
102. Tat'yana Nagnibeda. An upper bound for the spectral radius of a random walk on surface groups. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 240(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 2):154–165, 293–294, 1997.
103. J. Nash. Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.*, 80:931–954, 1958.
104. C. St. J. A. Nash-Williams. Random walk and electric currents in networks. *Proc. Cambridge Philos. Soc.*, 55:181–194, 1959.
105. V. I. Oseledec. A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. *Trudy Moskov. Mat. Obšč.*, 19:179–210, 1968.
106. Alan L. T. Paterson. *Amenability*, volume 29 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1988.
107. Rémi Peyre. A probabilistic approach to Carne's bound. *Potential Anal.*, 29(1):17–36, 2008.
108. J. W. Pitman. On coupling of Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(4):315–322, 1976.
109. Christophe Pittet and Laurent Saloff-Coste. A survey on the relationships between volume growth, isoperimetry, and the behavior of simple random walk on cayley graphs, with examples. 2001,unpublished.
110. Georg Pólya. Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. *Math. Ann.*, 84(1-2):149–160, 1921.
111. H. L. Royden. *Real Analysis*. Macmillan Publishing Company, New York, third edition, 1988.
112. Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
113. David Ruelle. Ergodic theory of differentiable dynamical systems. *Publications Mathématiques de l'IHÉS*, 50:27–58, 1979.
114. Mikko Salo. Lecture notes on Fourier analysis and distribution theory. http://users.jyu.fi/~salomi/lecturenotes/FA_distributions.pdf.
115. Caroline Series. Martin boundaries of random walks on Fuchsian groups. *Israel J. Math.*, 44(3):221–242, 1983.
116. Yehuda Shalom and Terence Tao. A finitary version of Gromov's polynomial growth theorem. *Geom. Funct. Anal.*, 20(6):1502–1547, 2010.
117. Frank Spitzer. *Principles of Random Walk*. Springer-Verlag, New York-Heidelberg, second edition, 1976. Graduate Texts in Mathematics, Vol. 34.

118. Richard P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997. With a foreword by Gian-Carlo Rota, Corrected reprint of the 1986 original.
119. J. Michael Steele. Kingman's subadditive ergodic theorem. *Ann. Inst. H. Poincaré Probab. Statist.*, 25(1):93–98, 1989.
120. Robert S. Strichartz. *A Guide to Distribution Theory and Fourier Transforms*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1994.
121. Terence Tao. *An introduction to measure theory*, volume 126 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2011.
122. Terry Tao. A proof of Gromov's theorem. https://terrytao.wordpress.com/2010/02/18/a-proof-of-gromovs-theorem/, Feb 2010.
123. Arkady Tempelman. *Ergodic theorems for group actions*, volume 78 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1992. Informational and thermodynamical aspects, Translated and revised from the 1986 Russian original.
124. Colin J. Thompson. *Mathematical statistical mechanics*. The Macmillan Co., New York; Collier-Macmillan Ltd., London, 1972. A Series of Books in Applied Mathematics.
125. J. Tits. Free subgroups in linear groups. *J. Algebra*, 20:250–270, 1972.
126. N. Th. Varopoulos. Convolution powers on locally compact groups. *Bull. Sci. Math. (2)*, 111(4):333–342, 1987.
127. Nicholas Th. Varopoulos. Chaînes de Markov et inégalités isopérimétriques. *C. R. Acad. Sci. Paris Sér. I Math.*, 298(18):465–468, 1984.
128. Nicholas Th. Varopoulos. Long range estimates for Markov chains. *Bull. Sci. Math. (2)*, 109(3):225–252, 1985.
129. Nicholas Th. Varopoulos. Théorie du potentiel sur des groupes et des variétés. *C. R. Acad. Sci. Paris Sér. I Math.*, 302(6):203–205, 1986.
130. Peter Walters. A dynamical proof of the multiplicative ergodic theorem. *Trans. Amer. Math. Soc.*, 335(1):245–257, 1993.
131. Wolfgang Woess. A local limit theorem for random walks on certain discrete groups. In *Probability measures on groups (Oberwolfach, 1981)*, volume 928 of *Lecture Notes in Math.*, pages 467–477. Springer, Berlin, 1982.
132. Wolfgang Woess. *Random Walks on Infinite Graphs and Groups*, volume 138 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2000.
133. Wolfgang Woess. Lamplighters, Diestel-Leader graphs, random walks, and harmonic functions. *Combin. Probab. Comput.*, 14(3):415–433, 2005.
134. Eberhard Zeidler. *Nonlinear functional analysis and its applications. I*. Springer-Verlag, New York, 1986. Fixed-point theorems, Translated from the German by Peter R. Wadsack.
135. Andrzej Żuk. A remark on the norm of a random walk on surface groups. *Colloq. Math.*, 72(1):195–206, 1997.

Index

S. P. Lalley, *Random Walks on Infinite Groups*, Graduate Texts in Mathematics 297,
https://doi.org/10.1007/978-3-031-25632-5